MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963
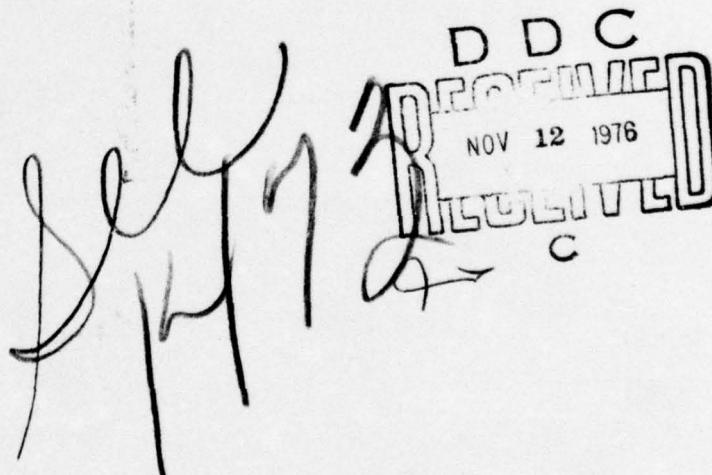
MRC Technical Summary Report #1655

ASYMPTOTIC THEORY FOR SOME FAMILIES
OF TWO-SAMPLE NONPARAMETRIC
STATISTICS

Lars Holst and J. S. Rao

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

July 1976

(Received April 19, 1976)

Approved for public release
Distribution unlimited

DDC

NOV 12 1976

RECEIVED
C

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ASYMPTOTIC THEORY FOR SOME FAMILIES OF
TWO-SAMPLE NONPARAMETRIC STATISTICS

Lars Holst[†] and J. S. Rao[‡]

Technical Summary Report #1655
July 1976

ABSTRACT

Let $X_1, \ldots, X_{m-1}$ and $Y_1, \ldots, Y_n$ be independent random samples from two
continuous distribution functions F and G respectively on the real line. We wish
to test the null hypothesis that these two parent populations are identical. Let
$X'_1 \leq \ldots \leq X'_{m-1}$ be the ordered X-observations. Denote by S_k the number of Y-
observations falling in the interval $[X'_{k-1}, X'_k)$, $k = 1, \ldots, m$. This paper studies
the asymptotic distribution theory and limiting efficiencies of families of test
statistics for the null hypothesis, based on these numbers $\{S_k\}$. Let $h(\cdot)$ and
$\{h_k(\cdot), k = 1, \ldots, m\}$ be real-valued functions satisfying some simple regularity
conditions. Asymptotic theory under the null hypothesis as well as under a suitable
sequence of alternatives, is studied for test statistics of the form $\sum_{k=1}^{m} h(S_k)$, based
symmetrically on S_k's and those of the form $\sum_{k=1}^{m} h_k(S_k)$ which are not symmetric in
$\{S_k\}$. It is shown here that tests of the symmetric type have poor asymptotic
performance in the sense that they can only distinguish alternatives at a "distance"
of $n^{-\frac{1}{4}}$ from the hypothesis. Among this class of symmetric tests, which includes
for instance the well known Run test and the Dixon test, it is shown that the Dixon
test has the maximum asymptotic relative efficiency. On the other hand, tests of
the nonsymmetric type can distinguish alternatives converging at the more standard
rate of $n^{-\frac{1}{2}}$. Wilcoxon-Mann-Whitney test is an example which belongs to this
class. After investigating the asymptotic theory under such alternatives, methods are
suggested which allow one to select an "optimal" test against any specific alternative,
from among tests of the type $\sum_{k=1}^{m} h_k(S_k)$. Connections with rank tests are briefly ex-
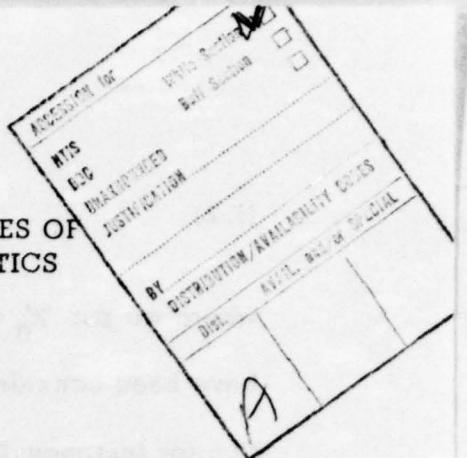plored and some illustrative examples provided.

† Mathematics Research Center, University of Wisconsin, Madison on leave from
   the Department of Mathematics, Uppsala University, Sweden.
‡ Department of Statistics, University of Wisconsin, Madison.

ASYMPTOTIC THEORY FOR SOME FAMILIES OF TWO-SAMPLE NONPARAMETRIC STATISTICS

Lars Holst[†] and J. S. Rao[‡]

1. Introduction and notations

Let $X_1, \ldots, X_{m-1}$ and $Y_1, \ldots, Y_n$ be independent random samples
from two populations with continuous distribution functions (d. f. s) $F(x)$
and $G(y)$ respectively. We wish to test if these two populations are
identical i.e., the hypothesis that the two d. f. s. are the same. A simple
probability integral transformation carrying $z \to I(z)$ would permit us to
assume that the support of both the probability distributions is the unit
interval $[0,1]$ and that the first of them is the uniform d. f. on $[0,1]$.
For the purposes of our discussion, this probability transformation can be
done without loss of any generality as will be apparent soon. Thus from
now on, we will assume that this reduction has been effected and that the
first sample is from the uniform distribution $U(0,1)$. Let $G^* = G \cdot F^{-1}$
denote the d. f. of the second sample after the probability transformation.
The null hypothesis we wish to test, specifies

$$(1.1) \qquad\qquad H_0: \; G^*(y) = y, \qquad 0 \leq y \leq 1 \; .$$

Let $0 \leq X'_1 \leq \ldots \leq X'_{m-1} \leq 1$ be the order statistics from the first
sample. The sample spacings $(D_1, \ldots, D_m)$ for the X-values are defined by

$$(1.2) \qquad D_k = X'_k - X'_{k-1}, \qquad k = 1, \ldots, m$$

where we put $X'_0 = 0$ and $X'_m = 1$. Tests based on these sample spacings have been considered in the literature for the goodness-of-fit problems. See for instance Darling (1953), Pyke (1965) and Rao and Sethuraman (1975). Define for $k = 1, \ldots, m$

$$(1.3) \qquad S_k = \text{number of } y_j\text{'s in the interval } [X'_{k-1}, X'_k).$$

Our aim is to study various test statistics based on these numbers $\{S_1, \ldots, S_m\}$ for testing H_0. Since the numbers $\{S_k\}$ as well as the statistics based on them remain invariant under probability transformations, there is no loss of generality in making such a transformation on the data, as was done earlier. It may be remarked here that we take $(m-1)$ instead of the usual m observations in the first sample now m, we will assume that this reduction has been effected and that the since this yields m numbers $\{S_1, \ldots, S_m\}$ instead of $(m+1)$, leading to slightly simpler notation.

Our aim is to study the asymptotic theory as m and n tend to infinity. We will do this through a nondecreasing sequence of positive integers $\{m_\nu\}$ and $\{n_\nu\}$. We will assume throughout, that as $\nu \to \infty$,

$$(1.4) \qquad m_\nu \to \infty, \quad n_\nu \to \infty \text{ and } m_\nu/n_\nu = r_\nu \to \rho, \quad 0 < \rho < \infty.$$

Note that $\{D_k\}$ defined in (1.2) depend on m_ν the number of X-values and it is more appropriate to label them as $\{D_{k\nu}\}$. Similarly the numbers $\{S_k\}$ defined in (1.3) depend on both m_ν and n_ν and should therefore

-2-

be denoted by $\{S_{k\nu}\}$. Thus we are dealing with triangular arrays of random variables $\{D_{k\nu}, \ k = 1, \ldots, m_\nu\}$ and $\{S_{k\nu}, \ k = 1, \ldots, m_\nu\}$ for $\nu \geq 1$. Corresponding to the νth $(\nu \geq 1)$ array, let $h_\nu(\cdot)$ and $\{h_{k\nu}(\cdot), \ k = 1, \ldots, m_\nu\}$ be real-valued functions satisfying certain regularity conditions (see Condition (A) of Section 2). We now define

(1.5)
$$T_\nu = \sum_{k=1}^{m_\nu} h_{k\nu}(S_{k\nu})$$

and

(1.6)
$$T_\nu^* = \sum_{k=1}^{m_\nu} h_\nu(S_{k\nu})$$

based on the $(m_\nu - 1)$ X-values and the n_ν Y-values. Though T_ν^* is a special case of T_ν when $\{h_{k\nu}(\cdot)\}$ do not depend on k, we will distinguish these two cases since their asymptotic behavior is quite different in the non-null situation. We may point out here that the Wald-Wolfowitz (1940) Run test and the Dixon (1940) test are of the form T_ν^* while the Wilcoxon-Mann-Whitney (1947) test is of the form T_ν. In fact, any linear function based on the X-ranks in the combined sample, can be expressed as a special case of T_ν. We will discuss more of this in Section 7.

A few words about the notations: Though the quantities m, n, r, D_k, S_k as well as the functions $h(\cdot)$, $\{h_k(\cdot)\}$ depend on ν, for notational convenience we shall suppress the suffix ν, except where it is essential. Thus for instance, $T_\nu = \sum_{k=1}^{m} h_k(S_k)$, $T_\nu^* = \sum_{k=1}^{m} h(S_k)$ and r will stand for (m/n) etc. We will also indicate the probability law of a random variable (or random vector) X by $\mathcal{L}(X)$.

-3-

$N(\mu, \Sigma)$ will represent a normal distribution with mean μ and covariance matrix Σ throughout while $N(0,0)$ stands for the degenerate distribution at the point zero. For $0 < x < \infty$, $Po(x)$ will represent the Poisson distribution with mean x and

$$(1.7) \qquad \pi_j(x) = e^{-x} \cdot x^j/j!, \qquad j = 0,1,2,\ldots$$

the Poisson probability of j. For $\underline{p} = (p_1, \ldots, p_m)$, $mult(n, \underline{p})$ will denote the m-dimensional multinomial distribution with n trials and cell probabilities $(p_1, \ldots, p_m)$. Let $\Gamma(a, b)$ denote the Gamma distribution with density

$$(1.8) \qquad e^{-x/b} \cdot x^{a-1}/b^a \cdot \Gamma(a) \text{ for } 0 < x < \infty \ .$$

A negative exponential i.e., a $\Gamma(1,1)$ random variable will be denoted throughout by W. η will stand for a geometric random variable with pdf

$$(1.9) \qquad P(\eta = j) = \rho/(1+\rho)^{j+1}, \qquad j = 0,1,2,\ldots$$

for $0 < \rho < \infty$.

Also for any random variable X_n, we write $X_n = o_p(g(n))$ if $X_n/g(n) \to 0$ in probability and we write $X_n = O_p(g(n))$ if for each $\epsilon > 0$, there is a $K_\epsilon < \infty$ such that $P\{|X_n/g(n)| > K_\epsilon\} < \epsilon$ for all n sufficiently large. Finally $[x]$ will denote the largest integer contained in x.

We shall consider a sequence of alternatives specified by the d. f. s.

$$(1.10) \qquad G_m^*(y) = y + (L_m(y))/m^\delta, \qquad 0 \le y \le 1$$

-4-

where $L_m(0) = L_m(1) = 0$ and $\delta \geq \frac{1}{4}$. In terms of the original d.f.s F and G, the null hypothesis specifies $G = F$, while under the alternatives we have a sequence of d.f.s G_m that converge to F as the sample size increases. Indeed

$$(1.11) \qquad L_m(y) = m^\delta (G_m(F^{-1}(y)) - y) .$$

We assume that there is a function $L(y)$ on $(0,1)$ to which $L_m(y)$ converges. For the other assumptions on $L_m(\cdot)$ and $L(\cdot)$ refer to Theorems 3.1 and 4.1. This sequence of alternatives (1.10) is smooth in a certain sense and has been considered before. See for instance Rao and Sethuraman (1975) or Holst (1972).

The organization of this paper is as follows: In Section 2, we establish some preliminary results. Theorem 2.1 gives asymptotic distribution of functions of multinomial frequencies while Theorem 2.2 establishes a result on the limit distributions of non-symmetric spacings statistics, which is of independent interest. These results are combined in Theorem 3.1 to obtain the limit distribution of T_ν under the alternatives (1.10) with $\delta = \frac{1}{2}$. It is clear that putting $L_m(y) \equiv 0$ in this theorem, gives the asymptotic distribution of T_ν under H_0. Section 4 deals with the symmetric statistics T_ν^*. Theorem 4.1 gives the asymptotic distribution of T_ν^* under the sequence of alternatives (1.10) with $\delta = \frac{1}{4}$. Again putting $L_m(y) \equiv 0$ gives the limit distribution of T_ν^* under H_0. It is interesting to note that the symmetric classes of tests T_ν^* can only distinguish alternatives

converging to the hypothesis at the slow rate of $n^{-\frac{1}{4}}$ unlike the non-symmetric statistics which can discriminate alternatives at the more usual distance of $n^{-\frac{1}{2}}$. Similar results hold for tests based on sample spacings depending on whether or not one considers symmetric statistics. See for instance Rao and Sethuraman (1969, 1975). Some further asymptotic theory of spacings statistics is currently under investigation by the present authors. Asymptotic efficiencies and applications are discussed separately for the non-symmetric and the symmetric cases, in Sections 5 and 6 respectively. Section 7 contains some further remarks and discussion.

2. Some preliminary results:

The following regularity conditions which limit the growth of the functions as well as supply certain smoothness properties, will be needed in connection with the results of this and the next section.

Condition (A): We will say that the real-valued functions $\{h_k(\cdot)\}$ defined on $\{0, 1, 2, \ldots\}$ satisfy Condition (A) if they are of the form

$$(2.1) \qquad h_k(j) = h(k/(m+1), j), \qquad k = 1, \ldots, m \qquad j = 0, 1, 2, \ldots$$

for some function $h(u, j)$ defined for $0 < u < 1$, $j = 0, 1, 2, \ldots$ with the properties

    (i) $h(u, j)$ is continuous in u except for finitely many u and the discontinuity set if any, does not depend on j.

    (ii) $h(u, j) \not\equiv c \cdot j + h(u)$ for some real number c.

    (iii) For some $\delta > 0$, there exist constants c_1 and c_2 such that

(2.2) $|h(u, j)| \leq c_1 \cdot [u(1-u)]^{-\frac{1}{2}+\delta} \cdot (j^{c_2}+1)$ for all $0 < u < 1$ and $j = 0, 1, 2, \ldots$

<u>Condition (A')</u>: We will say that the real-valued functions $\{g_k(\cdot)\}$ defined on $[0, \infty)$ satisfy Condition (A') if they are of the form

$$g_k(x) = g(k/(m+1), x), \quad k = 1, \ldots, m \text{ and } 0 \leq x < \infty$$

for some function $g(u, x)$ defined for $0 < u < 1$ and $0 \leq x < \infty$ with the properties

    (i) $g(u, x)$ is continuous in u except for finitely many u and the discontinuity set if any, does not depend on x.

    (ii) $g(u, x) \not\equiv c \cdot x + g(u)$ for some real number c.

and (iii) for some $\delta > 0$, there exist constants c_1 and c_2 such that

(2.3) $\qquad |g(u, x)| \leq c_1 \cdot [u(1-u)]^{-\frac{1}{2}+\delta} \cdot (x^{c_2}+1)$ for all $0 < u < 1$ and $0 \leq x < \infty$.

We require the following simple lemma.

<u>Lemma 2.1.</u> Let $h(u)$ defined for $0 < u < 1$, be continuous except for finitely many u and have the property: for some $a > -1$, there exists a constant c such that

(2.4) $\qquad\qquad\qquad |h(u)| \leq c \cdot [u(1-u)]^a$.

    Then

$$(1/m)\sum_{k=1}^{m} h(k/(m+1)) \to \int_0^1 h(u)du \text{ as } m \to \infty.$$

<u>Proof:</u> Define the step function

-7-

$$h_m(u) = h(([mu]+1)/(m+1)), \quad 0 < u < 1 \; .$$

Then clearly $h_m(u) \to h(u)$ as $m \to \infty$ except for finitely many points.
Observe

$$\left| h_m(u) \right| \leq 2c[u(1-u)]^a$$

for m sufficiently large and $\int_0^1 [u(1-u)]^a < \infty$ for $a > -1$. Thus by the Lebesgue Dominated Convergence theorem,

$$(1/m) \sum_{k=1}^m h(k/(m+1)) = \int_0^1 h_m(u)du \to \int_0^1 h(u)du$$

which proves the assertion. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Turning to the main problem, we will obtain the distribution of T_ν defined in (1.5) essentially in two steps. First we consider the statistic T_ν for given values of the X-spacings $\underline{\underline{D}} = \{D_1, \ldots, D_m\}$. Since the numbers $\{S_1, \ldots, S_m\}$ given $\underline{\underline{D}}$ have a multinomial distribution, we need a result on the multinomial sums. We formulate this part of the result in Theorem 2.1. The expressions for the asymptotic mean and variance of this conditional distribution of T_ν given $\underline{\underline{D}}$, are functions of $\underline{\underline{D}}$. In Theorem 2.2, we formulate a general result on the limit distributions of functions of spacings, which allows us to handle in particular, these expressions for the asymptotic mean and variance. Theorem 3.1 of the next section combines these results along with other lemmas given there, thus giving the required asymptotic distribution of T_ν .

-8-

It is clear that the conditional distribution of the vector of occupancy numbers $\underline{S} = (S_1, \ldots, S_m)$ given the spacings vector $\underline{D} = (D_1, \ldots, D_m)$ is $\text{mult}(n, D_1, \ldots, D_m)$. Therefore the test statistic T_ν, conditional on $\underline{D}$, has under the null hypothesis the same distribution as the random variable

$$(2.5) \qquad Z_\nu = \sum_{k=1}^{m_\nu} h_k(\varphi_k)$$

where $(\varphi_1, \ldots, \varphi_m)$ is $\text{Mult}(n, p_1, \ldots, p_m)$ with cell probabilities $(p_1, \ldots, p_m)$ being equal to the spacings $(D_1, \ldots, D_m)$. Since the asymptotic mean and variance of Z_ν can be more simply stated in terms of Poisson random variables, we introduce a triangular array of independent Poisson random variables $\{\xi_{1\nu}, \ldots, \xi_{m_\nu \nu}\}$, $\nu \geq 1$ where $\xi_{k\nu}$ is $\text{Po}(n_\nu p_{k\nu})$ and set

$$(2.6) \qquad \lambda_\nu = \sum_{k=1}^{m} h_k(\xi_{k\nu})$$

$$\mu_\nu = E(\lambda_\nu), \quad \sigma_\nu^2 = \text{Var}(\lambda_\nu)$$

(2.7)

The following theorem on the asymptotic distribution of the multinomial sum Z_ν is due to Holst (1976a).

<u>Theorem 2.1.</u> Let $(\varphi_1, \ldots, \varphi_m)$ be $\text{mult}(n, p_1, \ldots, p_m)$ and Z_ν, μ_ν, and σ_ν be as defined in (2.5), (2.6) and (2.7). For $0 < q \leq 1$, set $M = [mq]$ and

-9-

$$(2.8) \qquad \lambda_{\nu q} = \sum_{k=1}^{M} h_k(\xi_k) \ .$$

Assume that there exists a $q_0 < 1$ such that

$$(2.9) \qquad \sum_{k=1}^{M} p_k \to P_q, \quad 0 < P_q < 1 \text{ for } q_0 \leq q < 1 \ ,$$

and

$$(2.10) \qquad \mathcal{L} \begin{pmatrix} (\lambda_{\nu q} - E\lambda_{\nu q})/m^{\frac{1}{2}} \\[2mm] \sum_{1}^{M} (\xi_k - np_k)/n^{\frac{1}{2}} \end{pmatrix} \to N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A_q & B_q \\ B_q & P_q \end{pmatrix} \right)$$

where A_q, B_q and P_q are such that as $q \to 1-$,

$$(2.11) \qquad A_q \to A_1, \ B_q \to B_1 \text{ and } P_q \to 1 \ .$$

Then as $\nu \to \infty$,

$$(2.12) \qquad \mathcal{L}((Z_\nu - \mu_\nu)/\sigma_\nu) \to N(0, \ A_1 - B_1^2) \ . \qquad \square$$

From (2.6) and (2.7) an explicit expression for the asymptotic mean is given by

$$(2.13) \qquad \mu_\nu = \sum_{k=1}^{m} \sum_{j=0}^{\infty} h_k(j) \ \pi_j(np_k)$$

using the notation (1.7). Under the null hypothesis, we have $p_k = D_k$, $k = 1, \ldots, m$ where $\underline{\underline{D}}$ are the spacings from $U(0,1)$. Thus we consider

-10-

$$(2.14) \qquad \mu(n\underline{D}) = \sum_{k=1}^{m} \sum_{j=0}^{\infty} h_k(j) \, \pi_j(nD_k) = \sum_{k=1}^{m} g_k(mD_k)$$

where

$$(2.15) \qquad g_k(x) = \sum_{j=0}^{\infty} h_k(j) \, \pi_j(x/r) \ .$$

Random variables of the type (2.14) have been considered by Darling (1953), Lecam (1958) and Pyke (1965). For the symmetric case i.e., when $g_k(x) = g(x)$ for all k , Darling (1953) obtained some limit theorems for certain special cases and Le Cam (1958) gave a complete characterization of the limit laws. See also Rao and Sethuraman (1975) for some results in the symmetric case and their asymptotic efficiencies. Pyke (1965) pointed out (cf. Section 6.2) that Le Cam's method could be used to study the non-symmetric case. Since no complete result or its proof is explicitly given there in a form useful for our purposes, we state and prove such a limit theorem (Theorem 2.2) for non-symmetric functions of uniform spacings. An especially useful form of Theorem 2.2 is given in Corollary 2.1.

The method of proof we adopt is a mixture of the methods used by Le Cam (1958) and Darling (1953) and this result on spacings is of some independent interest. Let $W_1, W_2, \ldots$ be independently and identically

-11-

distributed (i.i.d.) $\Gamma(1, 1)$ random variables i.e., with a negative exponential

distribution $(1-e^{-w})$, $w \geq 0$. Let $\{g_k(\cdot), \; k = 1, \ldots, m\}$ be real-

valued measurable functions. For $0 < q \leq 1$, let $M_\nu = [m_\nu \cdot q]$. Define

(2.16)
$$G_{q\nu} = \sum_{k=1}^{M} g_k(W_k) \; .$$

Theorem 2.2: Assume that the variance of $G_{q\nu}$ exists and is positive

for all q and ν i.e.,

(2.17)
$$0 < \mathrm{Var}(G_{q\nu}) = \sigma^2(G_{q\nu}) < \infty \quad \text{for all } q \text{ and } \nu \; .$$

Assume further that for each $q \in (0, 1]$

(2.18)
$$\mathcal{L}\left(\begin{matrix} (G_{q\nu} - EG_{q\nu})/\sigma(G_{1\nu}) \\ \sum_{1}^{M} (W_k - 1)/m^{\frac{1}{2}} \end{matrix} \right) \to N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} A_q & B_q \\ B_q & q \end{pmatrix} \right)$$

with A_q and B_q such that

(2.19)
$$A_q \to A_1 = 1 \quad \text{as} \quad q \to 1- $$

(2.20)
$$B_q \to B_1 \quad \text{as} \quad q \to 1- \; .$$

Then, as $\nu \to \infty$,

(2.21)
$$\mathcal{L}\left((\sum_{k=1}^{m} g_k(mD_k) - EG_{1\nu})/\sigma(G_{1\nu}) \right) \to N(0, 1-B_1^2) \; .$$

where $D_1, \ldots, D_m$ are spacings from $U(0, 1)$.

Proof: Choose $0 < q < 1$. For $z > 0$ consider i.i.d. random variables

$V_1, \ldots, V_m$ with $\Gamma(1, 1/z)$ distribution i.e., exponential with mean $1/z$.

-12-

The sum $(V_1 + \dots + V_m)$ has $\Gamma(m, 1/z)$ distribution, whose density is denoted by $f(\cdot)$ below. Using ideas of conditional expectation, the characteristic function of $\sum\limits_{k=1}^{M} g_k(mV_k)$ at u can be written as

(2.22)
$$G_M(z, u) = E(\exp(iu \sum_1^M g_k(mV_k)))$$

$$= \int_0^\infty E\{\exp(iu \sum_1^M g_k(m V_k)) \mid \sum_1^M V_k = t)\} f(t)dt$$

$$= \int_0^\infty \{\phi(t, u)\} z^m t^{m-1} e^{-zt}/(m-1)! \, dt, \quad \text{say.}$$

Since the conditional distribution of $(V_1, \dots, V_m)$ given $\sum\limits_1^m V_k = t$ is the same as that of $(tD_1, \dots, tD_m)$ which does not depend on z, it follows that neither does $\phi(t, u)$. Observe that

(2.23)
$$\phi(u) = \phi(1, u) = E(\exp(iu \sum_1^M g_k(mV_k)) \mid \sum_1^m V_k = 1)$$
$$= E(\exp(iu \sum_1^M g_k(m D_k))) \ .$$

On the other hand the function defined by

$$G_M^*(z, u) = \int_0^\infty \dots \int_0^\infty \exp(iu \sum_1^M g_k(mv_k))z^M \exp(-z \sum_1^M v_k) \ dv_1 \dots dv_M$$

is analytic in z for all complex z such that $\text{Re}(z) > 0$. This

function coincides with $G_M(z, u)$ on the positive real axis and hence for all z in the right half plane. Consider

$$G_M^*(m+iy, u) = (m+iy)^M \int_0^\infty \ldots \int_0^\infty \exp(iu \sum_1^M g_k(mv_k) - iy \sum_1^M v_k)$$

$$\exp(-\sum_1^M mv_k) dv_1 \ldots dv_M$$

(2.24)

$$= (1+iy/m)^M \exp(-iyM/m) \cdot$$

$$E(\exp(iu \sum_1^M g_k(W_k - iy \sum_1^M (W_k-1)/m)))$$

where $W_1, \ldots, W_M$ are i.i.d. $\Gamma(1, 1)$ random variables. It is easily checked that the conditions for using the complex Laplace inversion formula (see for instance Courant and Hilbert (1962) p. 536) on (2.22) are satisfied. On doing this inversion, we obtain for $c > 0$

$$\phi(t, u) \cdot t^{m-1} = (m-1)! \, (1/2\pi i) \int_{c-i\infty}^{c+i\infty} e^{zt} z^{-m} G_M^*(z, u) dz \ .$$

Putting $c = m$ and $t = 1$ in this formula and using (2.23) and (2.24) we get

$$\phi(u) = E(\exp(iu \sum_1^M g_k(mD_k)))$$

(2.25)
$$= m!/(2\pi m) \int_{-\infty}^\infty e^{m+iy} (m+iy)^{-m} (1+iy/m)^M \exp(-iyM/m)$$

$$E(\exp(iuG_q - iy \sum_1^M (W_k-1)/m))dy \ .$$

-14-

Putting $x = y/m^{\frac{1}{2}}$ and using Stirling's formula for $m!$ we obtain

$$\phi(u) = e^{o(1)} (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp(ix(m-M)/m^{\frac{1}{2}}) \cdot (1+ix/m^{\frac{1}{2}})^{M-m}$$

$$E(\exp(iuG_q - ix \sum_{1}^{M} (W_k-1)/m^{\frac{1}{2}})dx .$$

From this, it follows that

$$\psi_\nu(u) = E(\exp(iu(\sum_{k=1}^{m_\nu} g_{k\nu}(m_\nu D_{k\nu}) - EG_{q\nu})/\sigma(G_{1\nu})))$$

$$(2.26) \qquad = e^{o(1)}(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp(ix(m - M)/m^{\frac{1}{2}})(1 + ix/m^{\frac{1}{2}})^{M-m}$$

$$E(\exp(iu(G_{q\nu} - EG_{q\nu})/\sigma(G_{1\nu}) - ix \sum_{1}^{M} (W_k - 1)/m^{\frac{1}{2}}))dx .$$

The integrand in (2.26) is dominated by

$$h_\nu(x) = (1 + x^2/m_\nu)^{(M_\nu - m_\nu)/2} \rightarrow h(x) = e^{-x^2(1-q)/2} \text{ as } \nu \rightarrow \infty$$

and it is easily verified that

$$\int_{-\infty}^{\infty} h_\nu(x)dx \rightarrow \int_{-\infty}^{\infty} h(x)dx \text{ as } \nu \rightarrow \infty .$$

Thus using the extended Lebesgue dominated convergence theorem (see for instance Rao (1973), p. 136), it follows from assumption (2.18) and the formula (2.26) that as $\nu \rightarrow \infty$,

$$\psi_\nu(u) \rightarrow (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp(-(1-q)x^2/2)$$

$$(2.27)$$

$$\exp(-(A_q u^2 + 2B_q ux + qx^2)/2)dx = \exp(-(A_q - B_q^2)u^2/2) .$$

-15-

The continuity theorem for characteristic functions gives

$$(2.28) \quad \mathcal{L}((\sum_{1}^{M} g_k(mD_k) - EG_{q\nu})/\sigma(G_{1\nu}))) \to \mathcal{L}(X_q) = N(0, A_q - B_q^2) \text{ as } \nu \to \infty .$$

Using an analogous derivation, we can show

$$(2.29) \quad \mathcal{L}(\sum_{M+1}^{m} (g_k(mD_k) - Eg_k(W_k))/\sigma(G_{1\nu})) \to \mathcal{L}(X'_q) = N(0, 1 - A_q - (B_1 - B_q)^2)$$

$$\text{as } \nu \to \infty .$$

Combining (2.28) and (2.29) with (2.19) and (2.20), we get

$$(2.30) \quad \mathcal{L}(X_q) \to N(0, 1 - B_1^2) \text{ as } q \to 1-$$

and

$$(2.31) \quad \mathcal{L}(X'_q) \to N(0, 0) \quad \text{as } q \to 1- .$$

Hence using the argument on pp. 13-14 of LeCam (1958) the assertion of the theorem follows. □

The following corollary gives a simple sufficient condition on the functions $g_k(\cdot)$ in order that Theorem (2.2) holds.

<u>Corollary 2.1.</u> The asymptotic normality asserted in Theorem 2.2 holds for any set of functions $\{g_k(\cdot)\}$ which satisfy condition (A') .

<u>Proof:</u> To prove this corollary, we need to check that the assumptions (2.17) to (2.20) hold when condition (A') is satisfied. It can be easily checked that if $g(u, x)$ satisfies condition (A'), then

$$\int_0^\infty g(u, x) e^{-x} dx, \quad \int_0^\infty g^2(u, x) e^{-x} dx \text{ as well as } \int_0^\infty g(u, x)(x-1) e^{-x} dx$$

-16-

satisfy conditions of Lemma 2.1 in u. Thus from the definition of $G_{q\nu}$ and Lemma 2.1, as $m \to \infty$,

$$(2.32) \qquad E(G_{q\nu})/m = (1/m) \sum_{k=1}^{[mq]} Eg(k/(m+1), W_k) \to \int_0^q Eg(u, W)du$$

$$(2.33) \qquad \text{Var}(G_{q\nu})/m = (1/m) \sum_{k=1}^{[mq]} \text{Var}(g(k/(m+1), W_k)) \to \int_0^q (\text{Var } g(u, W))du$$

and

$$(2.34) \qquad \text{Cov}(G_{q\nu}, \sum_1^M W_k)/m = (1/m) \sum_{k=1}^{[mq]} \text{Cov}(g(k/(m+1), W_k), W_k)$$

$$\to \int_0^q \text{Cov}(g(u, W), W)du$$

where W is a $\Gamma(1,1)$ random variable. Again from (2.3) of condition (A'), all these limits are finite. These are also continuous in q so that (2.19) and (2.20) are satisfied.

Finally to check the asymptotic normality in (2.18) or equivalently of

$$(2.35) \qquad \sum_{k=1}^{[mq]} \{a(g_k(W_k) - Eg_k(W_k)) + (W_k-1)\} = \sum_{k=1}^{[mq]} g_k^*(W_k), \quad \text{say}$$

for all real a, we have only to verify the Lindeberg condition for the non-identical case. It is easily seen that if $\{g_k(\cdot)\}$ satisfy condition (A'), so do $\{g_k^*(\cdot)\}$ defined in (2.35). Let

$$(2.36) \qquad \sigma_k^2 = Eg_k^{*2}(W_k) \quad \text{and} \quad s_{[mq]}^2 = \sum_{k=1}^{[mq]} \sigma_k^2 .$$

-17-

Since $\{g_k^*(\cdot)\}$ satisfy condition (A'), we have as in (2.33) that

$$(2.37) \qquad s_{[mq]}^2/m = (1/m) \sum_{k=1}^{[mq]} \int_0^\infty g_k^{*2}(w) e^{-w} dw$$

converges to a finite non-zero constant from Lemma 2.1. Now consider

$$(1/s_{[mq]}^2) \sum_{k=1}^{[mq]} \int_{|x| > \epsilon \, s_{[mq]}} g_k^{*2}(x) e^{-x} dx$$

$$\leq (m/s_{[mq]}^2) \cdot (1/m) \sum_{k=1}^{[mq]} \int_{|x| > \epsilon \, s_{[mq]}} c_1 [(k/(m+1))(1-k/(m+1)]^{-1+2\delta} \cdot$$

$$\cdot (x^{c_2}+1)^2 e^{-x} dx$$

$$\leq \{mc_1^2/s_{[mq]}^2\} \{(1/m) \sum_{k=1}^{[mq]} [(k/(m+1))(1-k/(m+1))]^{-1+2\delta}\}$$

$$\{\int_{|x| > \epsilon \, s_{[mq]}} (x^{c_2}+1)^2 e^{-x} dx\}$$

as $m \to \infty$, the quantities in the first two parentheses remain bounded because of (2.37) and Lemma 2.1 while the integral in the third parenthesis goes to zero for any $\epsilon > 0$ since $s_{[mq]}$ is of order $(\sqrt{m})$ from (2.37). Thus the Lindeberg condition is satisfied for (2.35) and thus the joint asymptotic normality required in (2.13) holds for functions satisfying condition (A'). □

3. Asymptotic distribution theory for nonsymmetric statistics

As mentioned in Section 1, η will denote a geometric random variable defined in (1.9) while W will represent a $\Gamma(1,1)$ random variable.

-18-

Where confusion is likely to arise, we will denote the expectations with respect to η and W by E_η and E_W respectively. Then it is easy to verify that for $j = 0, 1, 2, \ldots$

$$(3.1) \qquad P(\eta = j) = \rho / (1+\rho)^{j+1} = E_W(\pi_j(W/\rho)) \ .$$

If $h(u, j)$ is a function satisfying condition (A), we define for later use, the following additional functions.

$$(3.2) \qquad g_1(u, x) = \sum_{j=0}^{\infty} h(u, j) \pi_j(x) \ ,$$

$$(3.3) \qquad g_2(u, x) = \sum_{j=0}^{\infty} h^2(u, j) \pi_j(x)$$

and

$$(3.4) \qquad g_3(u, x) = \sum_{j=0}^{\infty} h(u, j)(j-x) \pi_j(x) \ .$$

When $h(u, j)$ satisfies condition (A), these are all well-defined and finite for all $\rho > 0$. Let

$$(3.5) \qquad H(u) = E_\eta h(u, \eta) = \sum_{j=0}^{\infty} h(u, j) \rho/(1+\rho)^{j+1}$$

$$= \sum_{j=0}^{\infty} h(u, j) E_W(\pi_j(W/\rho)) = E_W g_1(u, W/\rho)$$

for $0 < u < 1$ and

$$(3.6) \qquad \sigma^2 = \int_0^1 Var(h(u, \eta))du - \left(\int_0^1 Cov(h(u, \eta), \eta)du\right)^2 / Var(\eta)$$

$$= \sum_{j=0}^{\infty} \int_0^1 (h(u, j) - H(u))^2 \, du \, \rho/(1+\rho)^{j+1}$$

$$- \left(\sum_{j=0}^{\infty} \int_0^1 (h(u, j) - H(u)) du \, j\rho/(1+\rho)^{j+1}\right)^2 \cdot \rho^2/(1+\rho) \ .$$

-19-

Also from the Cauchy-Schwarz inequality

$$\left(\int_0^1 \mathrm{Cov}(h(u,\eta),\eta)\,du\right)^2 \le \left(\int_0^1 (\mathrm{Var}\,h(u,\eta))^{\frac{1}{2}}(\mathrm{Var}\,\eta)^{\frac{1}{2}}\,du\right)^2$$

$$\le \left(\int_0^1 \mathrm{Var}\,h(u,\eta)\,du\right)\left(\int_0^1 \mathrm{Var}(\eta)\,du\right)$$

$$= \mathrm{Var}(\eta)\cdot\left(\int_0^1 \mathrm{Var}\,h(u,\eta)\,du\right)$$

with equality if and only if $h(u,j) = c\cdot j + h(u)$ for some real number c and some function $h(u)$. Thus $\sigma^2 > 0$ for any function $h(u,j)$ satisfying condition (A). For $\underline{x} = (x_1,\ldots,x_m)$, we also define

$$(3.7) \qquad \mu_\nu(\underline{x}) = \sum_{k=1}^m g_1(k/(m+1),\,x_k) = \sum_{k=1}^m \sum_{j=0}^\infty h_k(j)\,\pi_j(x_k)$$

and write $\underline{W} = (W_1,\ldots,W_m)$ where the components are i.i.d. $\Gamma(1,1)$ random variables.

Before we proceed to state the theorem which gives the asymptotic distribution of T_ν under the alternatives, a few words about the sequence of alternatives. Consider the Y-observations from the distribution function given in (1.10), (1.11) with $\delta = \frac{1}{2}$ i.e.,

$$(3.8) \qquad A_m^{(1)}:\ G_m^*(y) = G_m(F^{-1}(y))$$

$$= y + L_m(y)/m^{\frac{1}{2}}, \qquad 0 \le y \le 1 .$$

-20-

<u>Assumption (B)</u>: We will assume that there exists a continuous function L(y) such that for $0 \leq y \leq 1$,

$$L_m(y) = m^{\frac{1}{2}}[G_m(F^{-1}(y)) - y] \rightarrow L(y) \text{ as } m \rightarrow \infty .$$

We also suppose that the derivatives $L_m'(y)$ and $L'(y) = \ell(y)$ exist and are continuous outside some fixed finite set $D \subset [0,1]$ and that finite left and right limits of the derivatives exist on the open interval $(0,1)$.

Given the X-sample, the probability of a Y-observation falling inside $[X_{k-1}', X_k')$, under the null hypothesis is given by the uniform spacings $\{D_k\}$. On the other hand, under the alternatives (3.8), this probability is given by

$$D_k^* = G_m(F^{-1}(U_k')) - G_m(F^{-1}(U_{k-1}'))$$

(3.9)
$$= D_k(1 + \Delta_k / m^{\frac{1}{2}})$$

where U_k', $k = 1, \ldots, m$ are order statistics from $U(0,1)$ with $U_0' = 0$, $U_m' = 1$ and
(3.10)
$$\Delta_k = [L_m(U_k') - L_m(U_{k-1}')]/D_k .$$

Note that $D_k > 0$ with probability one so that Δ_k is a well-defined random variable. We now state the main theorem of this section, whose proof will be completed in Lemmas 3.1 to 3.7. The conditions of this theorem are not in the most general form but are adequate to cover all the cases of interest discribed in Section 5.

-21-

<u>Theorem 3.1.</u> Let

(3.11) $$V_\nu = \sum_{k=1}^{m} (h(k/(m+1), S_k) - H(k/(m+1))) / m^{\frac{1}{2}} \cdot \sigma$$

with $H(u)$ and σ defined in (3.5) and (3.6) respectively. Assume that

(3.12) $$|h(u, j)| \leq c_1 (u(1-u))^\beta (1 + j^{c_2}) .$$

In addition to Assumption (B) assume that for some small $\epsilon > 0$,

(3.13) $$|L_m(t) - L_m(s)| \leq c_3 (t^\alpha - s^\alpha) \text{ for } 0 \leq s \leq t < \epsilon \text{ and for } (1-\epsilon) \leq s \leq t \leq 1$$

where $7/8 < \alpha < 1$, $\beta > -\frac{1}{2}$ and $(\alpha + \beta) > 1$. Then under the alternatives

(3.8),

(3.14) $$\mathcal{L}(V_\nu) \rightarrow N(b, I) ,$$

where

$$b = \int_0^1 \mathrm{Cov}(h(u, \eta), \eta) \, \ell(u) du \cdot \rho/(1+\rho)\sigma .$$

<u>Proof</u>: As explained in Section 2, the vector $(S_1, \ldots, S_m)$ given $\underline{\underline{D}}^*$ is

mult$(n, \underline{\underline{D}}^*)$ where the m-vector $\underline{\underline{D}}^*$ has the components D_k^* given in (3.9).

Using conditional expectations, we may write

(3.15)
$$E(e^{it\sigma V_\nu}) = E_{\underline{\underline{D}}^*} E(e^{it\sigma V_\nu} | \underline{\underline{D}}^*)$$

$$= E_{\underline{\underline{D}}^*} (J_\nu(\underline{\underline{D}}^*) K_\nu(\underline{\underline{D}}^*))$$

-22-

where

$$(3.16) \qquad J_\nu(\underline{\underline{D}}^*) = \exp(itm^{-\frac{1}{2}}[\mu(n\underline{\underline{D}}^*) - E\mu(\underline{\underline{W}}/\rho)])$$

(recall the definition (3.7) of μ_ν and the relation (3.5)) and

$$(3.17) \qquad K_\nu(\underline{\underline{D}}^*) = E(\exp(itm^{-\frac{1}{2}}[\sum_{k=1}^{m} h_k(S_k) - \mu(n\underline{\underline{D}}^*)]) \,\big|\, \underline{\underline{D}}^*) \ .$$

Now from Lemma 3.4, it follows that

$$\underset{\underline{\underline{D}}^*}{E}(J_\nu(\underline{\underline{D}}^*)) \to \exp(ibt - ct^2/2)$$

with b and c defined in (3.38) and (3.39) respectively. Hence by the continuity theorem for characteristic functions

$$(3.18) \qquad \mathcal{L}(m^{-\frac{1}{2}}[\mu_\nu(n\underline{\underline{D}}^*) - E\mu(\underline{\underline{W}}/\rho)]) \to N(b,c)$$

so that $J_\nu(\underline{\underline{D}}^*)$ converges in distribution. By Lemma 3.5, with probability one i.e., for almost every random vector $\underline{\underline{D}}^*$,

$$(3.19) \qquad K_\nu(\underline{\underline{D}}^*) \to e^{-dt^2/2}$$

with d as defined in (3.44). Combining (3.18) and (3.19), with probability one, the product $J_\nu(\underline{\underline{D}}^*) K_\nu(\underline{\underline{D}}^*)$ converges in distribution. But since $|J_\nu(\underline{\underline{D}}^*) K_\nu(\underline{\underline{D}}^*)| \le 1$, this also implies the convergence of the moments so that

$$(3.20) \qquad \underset{\underline{\underline{D}}^*}{E}(J_\nu(\underline{\underline{D}}^*) K_\nu(\underline{\underline{D}}^*)) \to \exp(ibt - (c+d)t^2/2) \ .$$

Using the continuity theorem for characteristic functions and Lemma 3.7, the assertion of the theorem follows. □

Lemma 3.1. If the conditions (3.12) and (3.13) of Theorem 3.1 hold, then

$$m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(k/(m+1), j) \left[\pi_j(nD_k^*) - \pi_j(nD_k) \right]$$

(3.21)

$$= m^{-1} \sum_{k=1}^{m} \Delta_k \sum_{j=0}^{\infty} h(k/(m+1), j) (j - nD_k) \pi_j(nD_k) + o_p(1)$$

where Δ_k is as defined in (3.10).

Proof: Applying the Cauchy-Schwartz inequality on the difference of the two sides in (3.21), we have

$$m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h_k(j) \left[\pi_j(nD_k^*) - \pi_j(nD_k) \{ 1 + (j - nD_k) \Delta_k / m^{\frac{1}{2}} \} \right]$$

$$\leq m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} |h_k(j)| \cdot \left| \exp \{ j \log(1 + \Delta_k / m^{\frac{1}{2}}) - nD_k \Delta_k / m^{\frac{1}{2}} \right.$$

(3.22)
$$\left. - 1 - (j - nD_k) \Delta_k / m^{\frac{1}{2}} \right| \pi_j(nD_k)$$

$$\leq m^{-\frac{1}{2}} \sum_{k=1}^{m} \left[\sum_{j=0}^{\infty} h_k^2(j) \pi_j(nD_k) \right]^{\frac{1}{2}} \left[\sum_{j=0}^{\infty} \exp \{ j \log(1 + \Delta_k / m^{\frac{1}{2}}) \right.$$

$$\left. - nD_k \Delta_k / m^{\frac{1}{2}} \} - 1 - (j - nD_k) \Delta_k / m^{\frac{1}{2}} \right|^2 \pi_j(nD_k) \right]^{\frac{1}{2}}.$$

After some elementary calculations, we see that the term in the second square brackets is

-24-

(3.23)
$$\exp(D_k \Delta_k^2 / r) - 1 - D_k \Delta_k^2 / r \; .$$

Since $h(u, x)$ satisfies condition (A'), using Theorem 2.2 and (3.23), it is clear that the right hand side of (3.22) can be estimated by

$$m^{-\frac{1}{2}} \sum_{k=1}^{m} c_1 \cdot g(k/m+1)(1+(mD_k)^{c_2}) O_p(D_k \Delta_k^2)$$

(3.24)

$$= \max_k (\Delta_k^2 / m^{\frac{1}{2}}) \cdot O_p(1) \; .$$

Now we show that this $\max(\Delta_k^2 / m^{\frac{1}{2}})$ goes to zero in probability when $\alpha > 7/8$. Observe that by (3.13)

$$|\Delta_k / m^{\frac{1}{4}}| = |L_m(U_k') - L_m(U_{k-1}')| / m^{\frac{1}{4}} \cdot D_k$$

(3.25)
$$\leq (U_k'^{\alpha} - U_{k-1}'^{\alpha}) / m^{\frac{1}{4}} \cdot D_k$$

$$\leq D_k^{\alpha} / m^{\frac{1}{4}} \cdot D_k$$

since $(t^{\alpha} - s^{\alpha}) \leq (t-s)^{\alpha}$ for $0 < s \leq t < 1$ and $\alpha < 1$. Also from Darling (1953) for any $\epsilon > 0$, we have

$$1 / \min_{1 \leq k \leq m} (m^{2+\epsilon} D_k) = O_p(1) \; .$$

Therefore from (3.25)

(3.26)
$$\max_{1 \leq k \leq m} |\Delta_k / m^{\frac{1}{4}}| \leq O_p(m^{(2+\epsilon)(1-\alpha) - \frac{1}{4}}) \; .$$

Since $\alpha > 7/8$, by choosing $0 < \epsilon < \frac{1}{4}(1-\alpha) - 2$, $\max|\Delta_k / m^{\frac{1}{4}}| \to 0$ in probability. This proves the lemma. $\qquad\square$

<u>Lemma 3.2.</u> If the conditions of Theorem 3.1 are satisfied then, for any $\epsilon > 0$,

$$(3.27) \quad \limsup_{m \to \infty} \left| m^{-1} \sum_{k=1}^{[m\epsilon]} \Delta_k \sum_{j=0}^{\infty} h(k/(m+1), j)(j - nD_k) \pi_j(nD_k) \right| \leq K\epsilon^{\alpha+\beta}$$

with probability one.

<u>Proof:</u> We have

$$\sum_{j=0}^{\infty} h_k(j)(j - nD_k)\, \pi_j(nD_k)$$

$$= (nD_k) \sum_{j=1}^{\infty} h_k(j)\, \pi_{j-1}(nD_k) - (nD_k) \sum_{j=0}^{\infty} h_k(j)\, \pi_j(nD_k)$$

$$= (nD_k) \sum_{j=0}^{\infty} [h_k(j+1) - h_k(j)]\, \pi_j(nD_k)$$

$$= nD_k\, g_4(k/(m+1), nD_k), \quad \text{say}$$

where

$$(3.28) \quad g_4(u, x) = \sum_{j=0}^{\infty} [h(u, j+1) - h(u, j)]\, \pi_j(x)$$

satisfies condition (A'). Hence the expression in (3.27)

$$m^{-1} \sum_{k=1}^{[m\epsilon]} \Delta_k \sum_{j=0}^{\infty} h(k/(m+1), j)(j - nD_k)\, \pi_j(nD_k)$$

$$(3.29)$$

$$= \sum_{k=1}^{[m\epsilon]} [L_m(U'_k) - L_m(U'_{k-1})]\, g_4(k/(m+1), nD_k) .$$

-26-

Using condition (3.13), and writing $M = [m\epsilon]$, this is

$$0 \le \Big| \sum_1^M [L_m(U_k') - L_m(U_{k-1}')] \, g_4(k/(m+1), nD_k) \Big|$$

(3.30)

$$\le c_1 c_3 \sum_1^M (U_k'^\alpha - U_{k-1}'^\alpha)(k/m)^\beta [1 + (nD_k)^{c_2}] \, .$$

We now make use of the representation of the spacings in terms of i.i.d. exponential r.v.s $W_1, W_2, \ldots$ with mean 1. Writing $\overline{W}_k = \sum_1^k W_j/k$, the RHS in (3.30) is

$$= C \cdot \frac{\epsilon^{\alpha+\beta}}{\overline{W}_m^{\alpha+c_2}} \sum_1^M [\overline{W}_k^\alpha (k/M)^\alpha - \overline{W}_{k-1}^\alpha ((k-1)/M)^\alpha](k/M)^\beta \cdot (W_k^{c_2} + \overline{W}_m^{c_2})$$

(3.31)

$$= C \cdot \epsilon^{\alpha+\beta} \cdot \overline{W}_m^{-(\alpha+c_2)} \cdot M^{-1} \sum_1^M \overline{W}_k^{\alpha-1} (k/M)^{\alpha+\beta-1} (W_k^{c_2+1} + W_k \cdot \overline{W}_m^{c_2})$$

$$\cdot [\{1 - (1 - W_k/k \cdot \overline{W}_k)^\alpha\}/(W_k/k\overline{W}_k)] \, .$$

Now by the strong law of large numbers $\overline{W}_k \to 1$ a.s. as $k \to \infty$ and hence

(3.32)
$$\{1 - (1 - W_k/k\overline{W}_k)^\alpha\}/(W_k/k\overline{W}_k) \to \alpha \quad \text{as } k \to \infty \, .$$

Using the Hölder inequality,

$$\frac{1}{M} \sum_1^M \overline{W}_k^{\alpha-1} (k/M)^{\alpha+\beta-1} W_k^{c_2+1} [\{1 - (1 - W_k/k\overline{W}_k)^\alpha\}/(W_k/k\overline{W}_k)]$$

(3.33)
$$\le \Big[\frac{1}{M} \sum_1^M \overline{W}_k^{(\alpha-1)p_1}\Big]^{1/p_1} \Big[\frac{1}{M} \sum_1^M (k/M)^{(\alpha+\beta-1)p_2}\Big]^{1/p_2}$$

$$\Big[\frac{1}{M} \sum_1^M W_k^{(c_2+1)p_3}\Big]^{1/p_3} \cdot \Big[\frac{1}{M} \sum_1^M \Big\{\frac{1 - (1 - W_k/k\overline{W}_k)^\alpha}{W_k/(k\overline{W}_k)}\Big\}^{p_4}\Big]^{1/p_4} \, .$$

-27-

Using the fact that if $a_k \to 1$ as $k \to \infty$, $n^{-1} \sum_1^n a_k \to 1$ as $n \to \infty$, the RHS in (3.33) converges a.s. to the finite limit

$$1 [\int_0^1 u^{(\alpha+\beta-1)p_2} du]^{1/p_2} [EW_2^{(c_2+1)p_3}]^{1/p_3} \alpha .$$

Similarly the other term involving $W_k \cdot \overline{W}_m^{c_2})$ in (3.31) can be handled so that we get the desired result. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 3.3.   Under the conditions of Theorem 3.1,

$$m^{-1} \sum_{k=1}^m \Delta_k \sum_{j=0}^\infty h(k/(m+1), j)(j - nD_k) \pi_j(nD_k)$$

$$\to \int_0^1 \ell(u) \, \mathrm{Cov}(h(u, \eta), \eta) du \cdot (\rho/1+\rho) \quad \text{in probability.}$$

Proof: As we have seen in (3.29),

$$m^{-1} \sum_{k=1}^m \Delta_k \sum_{j=0}^\infty h(k/(m+1), j)(j - nD_k) \pi_j(n D_k)$$

(3.34)

$$= m^{-1} \sum_{k=1}^m \Delta_k(nD_k) g_4(k/(m+1), nD_k)$$

where $g_4(u, x)$ is as defined in (3.28). For any fixed $\epsilon > 0$, we may consider the sum in (3.34) as consisting of 3 parts viz. $\sum_{k=1}^{[m\epsilon]}$, $\sum_{k=[m\epsilon]}^{[m(1-\epsilon)]}$ and $\sum_{k=[m(1-\epsilon)]}^m$. Lemma 3.2 shows that the first sum is negligible. A similar analysis can be used to demonstrate that the third term is also bounded a.s. by $K \epsilon^{\alpha+\beta}$.

Now it remains for us to show that in probability

-28-

$$m^{-1} \sum_{k=[m\epsilon]}^{[m(1-\epsilon)]} \Delta_k (nD_k) \, g_4(k/(m+1), \, nD_k) \to \int_\epsilon^{1-\epsilon} \ell(u) \cdot E(g_4(u, W/\rho)W/\rho)du$$

$$= \int_\epsilon^{1-\epsilon} \ell(u) \, E(\sum_{j=0}^\infty h(u,j)(j-W/\rho) \, \pi_j(W/\rho))du$$

(3. 35)

$$= \int_\epsilon^{1-\epsilon} \ell(u) \sum_{j=0}^\infty h(u,j)(j-(j+1)/(1+\rho)) \, P(\eta=j)du$$

$$= \int_\epsilon^{1-\epsilon} \ell(u) \cdot \text{Cov}(h(u,\eta), \eta)du \; \rho/(1+\rho)$$

The proof of Lemma 3. 3 will then be completed since we can choose ϵ arbitrarily small.

By our assumption $L'_{in}(y) = \ell_m(y)$ exists and is continuous except possibly for a finite number of points on $(\epsilon, 1-\epsilon)$. If $\ell_m(y)$ is continuous, then boundedness of $\ell_m(y)$ along with the fact $[xg_4(u,x)]$ satisfies condition (A') allows us to apply Theorem 2. 2 as follows: From the Glivenko-Cantelli theorem,

$$\max_k \left| \Delta_k - \ell_m(k/m+1) \right| \to 0 \qquad \text{with probability 1 .}$$

Also from Theorem 2. 2,

$$m^{-1} \sum_{k=[m\epsilon]}^{[m(1-\epsilon)]} nD_k \cdot g_4(k/(m+1), \, nD_k) = O_p(1) \; .$$

Hence the sum in (3. 35) has the same probability limit as

(3. 36)
$$m^{-1} \sum_{k=[m\epsilon]}^{[m(1-\epsilon)]} \ell_m(k/(m+1)) \cdot nD_k \cdot g_4(k/(m+1), \, nD_k)$$

which from Theorem 2. 2 is the required limit given in (3. 35).

Now if $\ell_m(\cdot)$ has a finite set of discontinuity points inside $(\epsilon, 1-\epsilon)$, this will not create any problems since the function is bounded in this interval.

-29-

Suppose now that $\ell_m(y)$ is continuous in $(0,1)$ except at $y = y_0$.
By our assumptions $\ell_m(y)$ has finite left and right limits at this point and
the point does not depend on m. Take $\delta > 0$ so that $0 < y_0 - \delta < y_0 + \delta < 1$.
From our assumptions and the Glivenko-Cantelli theorem it follows that with
probability one $|\Delta_k|$ is bounded whenever $|k/m - y_0| < \delta$ and m is
sufficiently large. From this it is easily seen by analogous arguments that
the contribution to the sum (3.35) from such terms in the neighborhood of
y_0 can be made arbitrary small by choosing δ sufficiently small. It is
obvious that the situation of a finite set of discontinuities of the first kind
can be handled the same way, if the discontinuity set does not depend on
m. This completes the proof of Lemma 3.3. □

Lemma 3.4. Let

$$J_\nu(\underline{D}^*) = \exp(itm^{-\frac{1}{2}}[\mu_\nu(n\underline{D}^*) - E\mu(\underline{W}/\rho)])$$

be as defined in (3.16). Then under the conditions of Theorem 3.1,

(3.37) $\qquad E(J_\nu(\underline{D}^*)) \to \exp(ibt - ct^2/2)$

where

(3.38) $\qquad b = \int_0^1 \mathrm{Cov}(h(u, \eta), \eta) \, \ell(u)du \, \rho/(1+\rho)$

and

(3.39) $\qquad c = \int_0^1 \mathrm{Var} \, g_1(u, W/\rho)du - (\int_0^1 \mathrm{Cov}(W, g_1(u, W/\rho))du)^2 \, .$

-30-

Proof: We can write

(3.40) $J_\nu(\underline{\underline{D}}^*) = \exp(itm^{-\frac{1}{2}}([\mu_\nu(n\underline{\underline{D}}^*) - \mu_\nu(n\underline{\underline{D}})] + [\mu_\nu(n\underline{\underline{D}}) - E\mu_\nu(W/\rho)]))$.

In Lemmas 3.1 to 3.3, we already established that the first part $m^{-\frac{1}{2}}[\mu_\nu(n\underline{\underline{D}}^*) - \mu_\nu(n\underline{\underline{D}})]$ converges in probability to b . Thus we need only show that

(3.41) $E(\exp(itm^{-\frac{1}{2}}[\mu_\nu(n\underline{\underline{D}}) - E\mu_\nu(W/\rho)])) \to \exp(-ct^2/2)$.

From the Condition (3.12) on $h(u, j)$, it follows using moments of the Poisson distribution, that

$$|g_1(u, x)| \leq \sum_{j=0}^{\infty} c_1(u(1-u))^\beta (1+x^{c_2}) \pi_j(x)$$

(3.42)

$$\leq c_1' (u(1-u))^\beta (1+x^{c_2'}) .$$

Thus $g_1(u, x)$ satisfies condition (A') . Hence Corollary 2.1 of Section 2 holds and the asymptotic normality of

$$\mu_\nu(n\underline{\underline{D}}) = \sum_{k=1}^{m_\nu} g_1(k/(m+1), nD_k)$$

is assured by Theorem 2.2. Further $Var(g_1(u, W/\rho))$ and $cov(w, g_1(u, W/\rho))$ as functions in u, satisfy the conditions of Lemma 2.1, so that as $\nu \to \infty$

$$m^{-1} Var(\sum_{k=1}^{m} g_1(k/(m+1), W_k/\rho) - m^{-2} Cov^2(\sum_1^m W_k, \sum_1^m g(k/(m+1), W_k/\rho))$$

(3.43) $\to \int_0^1 Var(g_1(u, W/\rho))du - (\int_0^1 Cov(W, g_1(u, W/\rho))du)^2$

$$= c .$$ $\square$

-31-

Lemma 3.5. Under the assumptions of Theorem 3.1, with probability one
i.e., for almost every $\underline{\underline{D}}^*$-vector,

$$K_\nu(\underline{\underline{D}}^*) = E(\exp(itm^{-\frac{1}{2}}[\sum_1^m h_k(S_k) - \mu(n\underline{\underline{D}}^*)])\,|\,\underline{\underline{D}}^*)$$

$$\rightarrow \exp(-dt^2/2) \text{ in probability}$$

where

$$(3.44) \qquad d = \int_0^1 E[g_2(u, W/\rho) - g_1(u, W/\rho)]^2\, du - \rho(\int_0^1 Eg_3(u, W/\rho)du)^2 .$$

Proof: The lemma will be proved by verifying that the conditions of

Theorem 2.1 hold and showing that $d = A_1 - B_1^2$. First we have by the

Glivenko-Cantelli theorem that with probability one

$$(3.45) \qquad \sum_{k=1}^M D_k^* = U'_M + m^{-\frac{1}{2}} L_m(U'_M) \rightarrow q = P_q$$

where $M = [mq]$ and U'_k is the kth order statistic from $U(0,1)$. Clearly

since $P_q = q \rightarrow 1$ as $q \rightarrow 1-$, conditions (2.9) and part of (2.11) of Theorem 2.1

hold.

For real numbers a and b, consider

$$(3.46) \qquad h_1(u, j) = ah(u, j) + bj .$$

It is easy to verify that if $h(u, j)$ satisfies condition (A), then so does

$h_1(u, j)$. Consider

-32-

$$(3.47) \qquad \zeta'_q = m^{-\frac{1}{2}} \sum_{k=1}^{M} (h_1(k/(m+1), \xi_k) - Eh_1(k/(m+1), \xi_k))$$

where $\xi_1, \ldots, \xi_m$ are independent and ξ_k is $P_0(nD_k^*)$. From the assumptions, it follows that for some positive constants $c_1, c_2, \ldots$ we have

$$V(\zeta'_q) = m^{-1} \sum_{k=1}^{M} Var(h_1(k/(m+1), \xi_k))$$

$$\leq m^{-1} c_1 \sum_{k=1}^{M} [(k/(m+1))(1 - k/(m+1))]^{\beta} ((nD_k^*)^{c_2} + 1)$$

(3.48)

$$\leq m^{-1} c_1' \sum_{k=1}^{m} [(k/(m+1))(1 - k/(m+1))]^{\beta} (nD_k^*)^{c_2} + c_3$$

$$\leq c_1'' (\sum_{k=1}^{m} (nD_k^*)^{c_4}/m)^{c_5} + c_3$$

by the Hölder inequality and Lemma 2.1. From the assumption (3.13),

$$nD_k^* = nD_k + n(L_m(U_k') - L_m(U_{k-1}'))m^{-\frac{1}{2}}$$

(3.49)

$$\leq nD_k + K_1 D_k m^{\frac{1}{2}} + K_2 D_k^{\alpha} m^{\frac{1}{2}}$$

$$\leq K_3(mD_k) + K_2(mD_k)^{\alpha} m^{\frac{1}{2}-\alpha} .$$

Using the representation $D_k = W_k / \sum_{1}^{m} W_k$ where $W_1, W_2, \ldots$ are i.i.d. $\Gamma(1,1)$ random variables, it follows by   law of large numbers that for $c > 0$,

-33-

$$\lim_{m \to \infty} m^{-1} \sum_{1}^{m} (mD_k)^C$$

is finite with probability one. As $\alpha > \frac{1}{2}$, we have, using the binomial theorem

$$m^{-1} \sum_{1}^{m} (mD_k^*)^{C_6} \leq m^{-1} K \sum_{1}^{m} (K_3 \, mD_k + K_2 (mD_k)^{\alpha} \, m^{\frac{1}{2}-\alpha} + 1)^{C_6+1} \to K_4$$

with probability one. Thus with probability one

(3.50) $$\lim \sup \text{Var}(\zeta_q') < \infty \ .$$

Now we will verify that

(3.51) $$\lim \inf \text{Var}(\zeta_q') > 0 \ .$$

By assumption (A), it follows that there exists an interval $[a, b] \subset (0, 1)$ and integers $j_1 \neq j_2$ such that $h_1(u, j_1) \neq h_1(u, j_2)$ for $a \leq u \leq b$. Again from the strong law of large numbers and our assumptions, it is easily seen that for any $0 < C < D < \infty$, with probability one

$$\# \{k; \ a < k/(m+1) < b, \quad C < nD_k^* < D\} /m \to K_1 > 0 \ .$$

Therefore for n sufficiently large,

$$\text{Var}(\zeta_q') \geq \sum_{\substack{a(m+1)<k<b(m+1)}} \text{Var}(h_1(k/(m+1), \xi_k))/m \geq K_2 > 0$$

with probability one. Hence (3.51) is satisfied with probability one. In a similar fashion it follows that

$$\limsup \sum_{k=1}^{m} E|h_1(k/(m+1), \xi_k)|^{2+\epsilon}/m < \infty .$$

Therefore the Liapunov condition

$$\sum_{1}^{M} E|h_1(k/(m+1), \xi_k)|^{2+\epsilon}/(m \, Var(\zeta'_q))^{1+\epsilon/2}$$

(3.52) $$= m^{-\epsilon/2} \sum_{1}^{M} E|h_1(k/(m+1), \xi_k)|^{2+\epsilon}/(Var(\zeta'_q))^{1+\epsilon/2} \cdot m$$

$$\rightarrow 0 \quad \text{as} \quad m \rightarrow \infty ,$$

is satisfied with probability one. Thus

$$\mathcal{L}(\zeta'_q/(Var(\zeta'_q))^{\frac{1}{2}}) \rightarrow N(0,1)$$

with probability one. By the next Lemma 3.6, we have that in probability

$$Var(\zeta'_q) \rightarrow a^2 A_q + 2ab B_q \rho^{-\frac{1}{2}} + b^2 q \rho^{-1})$$

where $A_q \rightarrow A_1$, $B_q \rightarrow B_1$ as $q \rightarrow 1-$. This verifies that the assumptions of Theorem 2.1 are satisfied with probability one. From the definition (3.44) of d as well as the expressions (3.56) and (3.57) for A_q and B_q, it follows

(3.54) $$d = A_1 - B_1^2$$

which proves the lemma. □

Lemma 3.6. Given $\underset{=}{D}^*$, let $(\xi_1, \ldots, \xi_m)$ be independent and ξ_k be $Po(nD_k^*)$. Under the assumptions of Theorem 3.1

(3.55) $m^{-1} \sum_{1}^{M} \text{Var}(h_1(k/(m+1), \xi_k)) \to a^2 A_q + 2ab B_q \rho^{-\frac{1}{2}} + b^2 q \rho^{-1}$

in probability where

(3.56) $$A_q = \int_0^q E[g_2(u, W/\rho) - g_1(u, W/\rho)]^2 \, du$$

and

(3.57) $$B_q = \rho^{-\frac{1}{2}} \int_0^q E(g_3(u, W/\rho)) \, du \; .$$

Proof: Recall from (3.30) that $h_1(u, j) = a h(u, j) + bj$. By calculations similar to those in Lemma 3.1, it follows that, for instance

$$m^{-1} \sum_{k=1}^{M} \sum_{j=0}^{\infty} h^2(k/(m+1), j) \left[\pi_j(n D_k^*) - \pi_j(n D_k) \right] \to 0$$

in probability.   Using Theorem 2.2, we get

$$m^{-1} \sum_{k=1}^{M} h^2(k/(m+1), j) \, \pi_j(n D_k) \to \int_0^q (E g_2(u, W/\rho)) du$$

in probability.   Therefore

$$m^{-1} \sum_{1}^{M} E h^2(k/(m+1), \xi_k) \to \int_0^q (E g_2(u, W/q)) du \; .$$

The other terms can be handled analogously which proves the assertion. $\square$

Lemma 3.7.

(3.58) $c + d = \sigma^2$

where c, d, σ^2 are defined in (3.39), (3.44) and (3.6) respectively.

Proof: Let W be a $\Gamma(1, 1)$ variable and let η' be a random variable such that η' given W is $\text{Po}(W/\rho)$. Then the unconditional distribution of η' is given by

-36-

$$P(\eta' = j) = \int_0^\infty e^{-W} \cdot (W/\rho)^j \, e^{-(W/\rho)}/j! \; dW$$

(3.59)
$$= \rho/(1+\rho)^{j+1}, \qquad j = 0, 1, 2, \ldots \quad .$$

Thus η' has the same distribution as the geometric random variable η defined in (1.9). Let E_W, V_W denote the expectation and variance over W while $E_{\eta|W}$ $V_{\eta|W}$ denote the conditional expectation and variance over η given W. Then from the definitions of g_1, g_2, g_3

(3.60) $\quad E_\eta h(u, \eta) = E_W E_{\eta|W} h(u, \eta) = E_W g_1(u, W/\rho)$

(3.61) $\quad E_\eta h^2(u, \eta) = E_W E_{\eta|W} h^2(u, \eta) = E_W g_2(u, W/\rho)$.

And after some elementary calculations,

$$\rho(1+\rho)^{-1} \, \text{Cov}(h(u, \eta), \eta) = E_W E_{\eta|W}[h(u, \eta)(\eta - W/\rho)]$$

(3.62)
$$= \text{Cov}(g_1(u, W/\rho), W) = E_W g_3(u, W/\rho).$$

Now from the definitions (3.39) and (3.44) of c and d and from identities (3.60), (3.61) and (3.62), we get

$$c + d = \int_0^1 \text{Var}(g_1(u, W/\rho))du - (\int_0^1 \text{Cov}(W, g_1(u, W/\rho))du)^2$$

$$+ \int_0^1 E[g_2(u, W/\rho) - g_1(u, W/\rho)]^2 du - \rho(\int_0^1 Eg_3(u, W/\rho)du)^2$$

$$= \int_0^1 V_W(E_\eta | W^{(h(u, \eta))})du + \int_0^1 E_W(V_\eta | W^{(h(u, \eta))})du$$

$$- (1+\rho) [\int_0^1 \text{Cov}(h(u, \eta), \eta) \, \rho/(1+\rho)du]^2$$

$$= \int_0^1 \text{Var}(h(u, \eta))du - [\int_0^1 \text{Cov}(h(u, \eta), \eta)du]^2 \rho^2/(1+\rho)$$

$$= \sigma^2 . \qquad \square$$

These lemmas 3.1 to 3.7 complete the proof of Theorem 3.1. The following lemma gives a simple sufficient condition for (3.13) hold.

Lemma 3.8. A sufficient condition for (3.13) to hold in a neighborhood of the origin is that

$$0 \le L'_m(u) \le c \cdot u^{\alpha - 1} \quad \text{for} \quad 0 < u < \epsilon .$$

Proof. We have for $0 < s < t < \epsilon$

$$0 \le \int_s^t (cu^{\alpha - 1} - L'_m(u))du$$

$$= c(t^\alpha - s^\alpha)/\alpha - (L_m(t) - L_m(s))$$

or

$$L_m(t) - L_m(s) \le c(t^\alpha - s^\alpha)/\alpha .$$

Since $L_m(0) = 0$ and $L'_m(u) \ge 0$, the assertion follows. $\qquad \square$

Corollary 3.1. Under the null hypothesis (1.1), the asymptotic distribution of V_ν defined in (3.11) is $N(0,1)$. $\square$

This result is a direct consequence of Theorem 3.1 and is obtained by taking $\ell(u) \equiv 0$, $0 < u < 1$ in (3.14). This corollary regarding the null distribution of V_ν can also be reformulated in the following interesting form using Lemma 2.1.

Corollary 3.1'. Let $\eta_1, \eta_2, \ldots$ be a sequence of i.i.d. geometric random variables with pdf given in (1.9). Then the asymptotic null distribution of $\sum_1^m h_k(S_k)$ is $N(E(\sum_1^m h_k(\eta_k)), \text{Var}(\sum_1^m h_k(\eta_k) - \beta \sum_1^m \eta_k))$ where β is the regression coefficient given by

$$\beta = \text{Cov}(\sum_1^m h_k(\eta_k), \sum_1^m \eta_k)/\text{Var}(\sum_1^m \eta_k)$$
$\square$

Remark: See also Holst (1976b) for a different proof of this result.

4. Asymptotic distribution theory for symmetric statistics

This section deals with the class of statistics symmetric in $\{S_1, \ldots, S_m\}$ i.e., statistics of the form

$$(4.1) \qquad T_\nu^* = \sum_{k=1}^{m_\nu} h(S_k)$$

for some given function $h(j)$. Clearly this class of symmetric statistics is also covered by the asymptotic theory discussed in the last section. Indeed if the function $h_k(j)$ does not vary with k i.e., the function $h(u, j)$ of the last section is a function only of j and is independent of u, then we obtain the symmetry in the numbers $\{S_1, \ldots, S_m\}$. But since $\int_0^1 \ell(u)du = 0$, it follows from Theorem 3.1 and Corollary 3.1 that the

-39-

asymptotic distribution of T_ν^* under the sequence of alternatives (3.8) coincides with that under the null hypothesis. Thus symmetric statistics of the type (4.1) can not distinguish alternatives that are at a 'distance' of $n^{-\frac{1}{2}}$ and have power zero against such close alternatives. Therefore in order to make efficiency comparisons, we have to consider the more distant alternatives with $\delta = \frac{1}{4}$ in (1.10). Let

$$(4.2) \qquad A_m^{(2)}: \quad G_m^*(y) = y + L_m^*(y)/m^{\frac{1}{4}}, \qquad 0 \le y \le 1$$

$$\text{with } L_m^*(u) = m^{\frac{1}{4}}(G_m(F^{-1}(u)) - u) \ .$$

For this symmetric situation, we will make the following slightly stronger assumptions:

__Assumption (B^*)__. Assume $L_m^*(u)$ is twice differentiable on $[0,1]$ and there is a function $L^*(u)$, $0 \le u \le 1$, which is twice continuously differentiable and such that

$$(4.3) \qquad L^*(0) = L^*(1) = 0, \quad \sup_{0 \le u \le 1} \left| L_m^{*''}(u) - L^{*''}o(u) \right| = o(1)$$

where $\ell^*(u)$, $\ell^{*'}(u)$ are the first and second derivatives of $L^*(u)$. Notice that for such smooth alternatives, the following also hold:

$$(4.4) \qquad \sup_{0 \le u \le 1} \left| L_m^*(u) - L^*(u) \right| = o(1), \qquad \sup_{0 \le u \le 1} \left| L_m^{*'}(u) - \ell^*(u) \right| = o(1) \ .$$

We define analogous to (3.9) and (3.10)

-40-

(4.5) $\qquad D_k^* = D_k(1 + \Delta_k^*/m^{\frac{1}{4}})$ with $\Delta_k^* = [L_m^*(U_k^{\shortmid}) - L_m^*(U_{k-1}^{\shortmid})]/D_k$.

We observe that under the above regularity conditions, we have

(4.6) $\qquad \max_{1 \le k \le m} |\Delta_k^*| \le \sup_{0 \le u \le 1} |\ell_m^*(u)| \le K < \infty$.

The following theorem gives the asymptotic distribution of the symmetric statistics T_ν^* under the alternatives (4.2). The proof of this theorem will follow Lemmas 4.1, 4.2 and 4.3.

<u>Theorem 4.1.</u> Suppose that there exist constants c_1 and c_2 such that

(4.7) $\qquad\qquad |h(j)| \le c_1(j^{c_2} + 1)$ for all j .

Let $L_m^*(u)$ satisfy Assumption (B^*) and let

(4.8) $\qquad\qquad V_\nu^* = \sum_{k=1}^m (h(S_k) - Eh(\eta))/m^{\frac{1}{2}} \sigma$

where

(4.9) $\qquad\qquad \sigma^2 = Var(h(\eta)) - [Cov(h(\eta), \eta)]^2/Var(\eta)$

and η is the geometric random variable defined in (1.9). Then under the alternatives (4.2)

(4.10) $\qquad\qquad \mathcal{L}(V_\nu^*) \to N(A, 1)$ as $\nu \to \infty$

where

(4.11) $\qquad\qquad A = (\int_0^1 \ell^2(u)du) \, Cov(h(\eta), \eta(\eta-1) - 4\eta/\rho) \, \rho^2/2(1+\rho)^2\sigma$. $\qquad \square$

-41-

<u>Lemma 4.1.</u>   Under the assumptions of Theorem 4.1, we have

$$m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \pi_j(nD_k) [(1 + \Delta_k/m^{\frac{1}{4}})^j \exp(-nD_k\Delta_k/m^{\frac{1}{4}})$$

(4.12) $$-1 - (j - nD_k)\Delta_k/m^{\frac{1}{4}} - \{j(j-1) - 2jnD_k + (nD_k)^2\}\Delta_k^2/m^{\frac{1}{2}} r]$$

$$= o_p(1) \ .$$

<u>Proof:</u>   Using the Cauchy-Schwarz inequality, we find that the difference in (4.12) can be estimated by

$$m^{-\frac{1}{2}} \sum_{k=1}^{m} c_1(1 + (nD_k)^{c_2}) [B(nD_k, \Delta_k/m^{\frac{1}{4}})]^{\frac{1}{2}}$$

where

$$B(x, y) = \sum_{j=0}^{\infty} \pi_j(x) [(1+y)^j \exp(-xy/r) - 1 - (j - x/r)y$$

$$- \{j(j-1) - 2jx/r + (x/r)^2\} y^2/2]^2$$

$$= \exp(xy^2/r) - 1 - xy^2/r - x^2y^4/2r^2 \ .$$

Therefore for small (xy^2), $B(x, y) = O((xy^2)^3)$. As $\max \Delta_k = O(1)$ and $\max(mD_k/\log m) = O_p(1)$, (cf. Darling (1953, p. 251)) we can estimate (4.12) by

(4.13) $$m^{-5/4} \sum_{k=1}^{m} g(nD_k)$$

for some function $g(\cdot)$ satisfying the conditions of Theorem 2.2. Therefore (4.12) is $O_p(m^{-\frac{1}{4}})$ which proves the lemma.   □

-42-

Lemma 4.2. Under the assumptions of Theorem 4.1, we have

$$(4.14) \qquad m^{-3/4} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \, \pi_j(n D_k)(j - n D_k) \Delta_k = o_p(1) \ .$$

Proof: From the assumed continuity of $\ell_m^*(u)$, it follows that for some $\tilde{U}_k$ where $U'_{k-1} \leq \tilde{U}_k \leq U'_k$ that

$$m^{-3/4} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \, \pi_j(n D_k)(j - n D_k) \Delta_k$$

$$(4.15)$$

$$= m^{-3/4} \sum_{k=1}^{m} \ell_m^*(\tilde{U}_k) \, g_3(n D_k)$$

where

$$(4.16) \qquad g_3(x) = \sum_{j=0}^{\infty} h(j) \, \pi_j(x) \, (j-x) \ .$$

Since $\int_0^1 \ell_m^*(u) du = 0$, we can choose θ_k such that $(k-1)/m < \theta_k < k/m$ and $\sum_{k=1}^{m} \ell_m^*(\theta_k) = 0$. Also since $\ell_m^{*'}$ is continuous, we can write for some $\tilde{\tilde{U}}_k$ such that $\min(\theta_k, \tilde{U}_k) \leq \tilde{\tilde{U}}_k \leq \max(\theta_k, \tilde{U}_k)$

$$m^{-3/4} \sum_{k=1}^{m} \ell_m^*(\tilde{U}_k) \, g_3(n D_k)$$

$$(4.17)$$

$$= m^{-3/4} \sum_{k=1}^{m} \ell_m^*(\theta_k) \, [g_3(n D_k) - E g_3(W)]$$

$$+ m^{-3/4} \sum_{k=1}^{m} \ell_m^{*'}(\tilde{\tilde{U}}_k) \, (\tilde{U}_k - \theta_k) \, g_3(n D_k) \ .$$

Now $\max \ell_m^{*'}(\tilde{\tilde{U}}_k) = O_p(1)$ and from the boundedness of the Kolmogorov-Smirnov statistic

$$\max_k \; m^{\frac{1}{2}} |\tilde{U}_k - \theta_k| = O_p(1) \; .$$

Using Theorem 2.2, we have

$$(4.18) \qquad m^{-3/4} \sum_{k=1}^m \ell_m^{*'}(\tilde{\tilde{U}}_k)(\tilde{U}_k - \theta_k) \, g_3(nD_k) = O_p(m^{-\frac{1}{4}}) \; .$$

Theorem 2.2 also gives

$$(4.19) \qquad m^{-1/2} \sum_{k=1}^m \ell_m^*(\theta_k) \, [g_3(nD_k) - Eg_3(W/\rho)] = O_p(1) \; .$$

Combining the estimates (4.18) and (4.19) in (4.17) yields the desired

result. □

Lemma 4.3. Under the assumptions of Theorem 4.1, we have

$$m^{-1} \sum_{k=1}^m \Delta_k^2 \sum_{j=1}^\infty h(j) \, \pi_j(nD_k) \, \{j(j-1) - 2jnD_k + (nD_k)^2\}/2$$

$$\to (\int_0^1 \ell_m^{*2}(u)du) \; \text{Cov}(h(\eta), \, \eta(\eta-1) - 4\eta/\rho) \cdot \; \rho^2/2(1+\rho)^2$$

in probability as $\nu \to \infty$.

Proof: Using the assumed smoothness of $\ell_m^*(u)$, $0 \le u \le 1$ and the

Glivenko-Cantelli theorem, it follows that

$$\max_{1 \le k \le m} |\Delta_k^2 - \ell_m^{*2}(k/m+1)| \to 0 \; \text{with Probability 1.}$$

And from Theorem 2.2, we have

$$m^{-1} \sum_{k=1}^m |\sum_{j=0}^\infty h(j) \, \pi_j(nD_k) \, \{j(j-1) - 2jmD_k/r + (mD_k/r)^2\}/2| = O_p(1) \; .$$

-44-

Hence it follows that it is sufficient to prove the stochastic convergence

of

(4.21)
$$m^{-1} \sum_{k=1}^{m} \ell_m^{*2}(k/(m+1)) \, g_5(nD_k)$$

where

(4.22)
$$g_5(x) = \sum_{j=0}^{\infty} h(j) \, \pi_j(x) \, \{j(j-1) - 2jx + x^2\}/2$$

to the limit in the assertion. Again Theorem 2.2 gives that

$$m^{-1} \sum_{k=1}^{m} \ell_m^{*2}(k/(m+1) \, g_5(nD_k) \rightarrow (\int_0^1 \ell^{*2}(u)du) \, Eg_5(W/\rho)$$

in probability. An elementary calculation shows that

$$Eg_5(W/\rho) = E \sum_{j=0}^{\infty} h(j) \, \pi_j(W/\rho) \, \{j(j-1) - 2j(W/\rho) + (W/\rho)^2\}/2$$

$$= \text{Cov}(h(\eta), \, \eta(\eta-1) - 4\eta/\rho) \cdot \rho^2/2(1+\rho)^2 \quad .$$

Thus the lemma is proved. □

Proof of Theorem 4.1. Following the method used in the proof of Theorem

3.1, it suffices to show that

$$m^{-\frac{1}{2}}[\mu_\nu(\, n\underline{\underline{D}}^*) - \mu_\nu(n\,\underline{\underline{D}})] \rightarrow A \quad \text{in probability.}$$

We have

$$m^{-\frac{1}{2}} [\mu_\nu(n \underline{\underline{D}}^*) - \mu_\nu(n \underline{\underline{D}})]$$

$$= m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) [\pi_j(n D_k^*) - \pi_j(n D_k)]$$

$$= m^{-\frac{1}{2}} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \pi_j(n D_k) [(1 + \Delta_k/m^{\frac{1}{4}})^j \exp(-n D_k \Delta_k/m^{\frac{1}{4}})$$

$$- 1 - j - n D_k \quad \Delta_k/m^{\frac{1}{4}} - \{j(j-1) - 2 j n D_k + (n D_k)^2\} \Delta_k^2/m^{\frac{1}{2}}$$

$$+ m^{-3/4} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \pi_j(n D_k)(j - n D_k) \Delta_k$$

$$+ m^{-1} \sum_{k=1}^{m} \sum_{j=0}^{\infty} h(j) \pi_j(n D_k) \{j(j-1) - 2 j n D_k + (n D_k)^2\} \Delta_k^2/2 .$$

Therefore the proof of Theorem 4.1 is complete from the preceding Lemmas 4.1, 4.2 and 4.3. $\qquad\square$

Taking $\ell^*(u) \equiv 0$, $0 < u < 1$ in Theorem 4.1 or putting $h_k(j) = h(j) \forall k$ in Corollaries 3.1, 3.1', we get the following result on the asymptotic null distribution for the symmetric statistics.

<u>Corollary 4.1.</u> Let V_ν^* be as defined in (4.8). Then under the null hypothesis (1.1) V_ν^* has asymptotically a $N(0,1)$ distribution if the function $h(\cdot)$ satisfies condition (4.7).

5. <u>Some further results and applications - the nonsymmetric case</u>

We have the sequence of alternatives given by (3.8) i.e.,

$$(5.1) \qquad G_m^*(y) = y + L_m(y)/m^{\frac{1}{2}}, \quad 0 < y < 1$$

-46-

where

$$L_m(u) = m^{\frac{1}{2}}(G_m(F^{-1}(u)) - u) \to L(u) \quad \text{as} \quad m \to \infty \ .$$

We now consider first the problem of finding the optimal choice of the function $h(u, j)$ for given alternatives $G_m^*(y)$. We use a technique analogous to that in Holst (1972).

<u>Theorem 5.1.</u>   If the sequence of alternatives is such that the assumptions of Theorem 3.1 are fulfilled, then an asymptotically most powerful (AMP) test of the hypothesis against the simple alternative (5.1) is to reject H_0 when

(5.2)
$$\sum_{k=1}^{m} \ell(k/m+1) S_k > c$$

where $\ell(u)$ is the derivative of $L(u)$, mentioned in (5.1). The asymptotic distribution of this optimal statistic is given by

(5.3)
$$\mathcal{L}(m^{-\frac{1}{2}} \sum_{k=1}^{m} \ell(k/(m+1))(S_k - 1/\rho)) \to N(0, \sigma^2)$$

under H_0 with

(5.4)
$$\sigma^2 = (\int_0^1 \ell^2(u)du) (1+\rho)/\rho^2$$

while under the alternatives (5.1)

(5.5)
$$\mathcal{L}(m^{-\frac{1}{2}} \sum_{k=1}^{m} \ell(k/(m+1)) (S_k - 1/\rho)) \to N(\rho^{-1}(\int_0^1 \ell^2(u)du \ , \ \sigma^2)$$

<u>Proof:</u>   From Theorem 3.1, it follows that the asymptotic power of the test which rejects H_0 when $\sum_{k=1}^{m} h(k/(m+1), S_k) > c$ is given by

-47-

$$P_h = \int_0^1 \text{Cov}(h(u, \eta), \eta) \, \ell(u) du / [\int_0^1 (\text{Var } h(u, \eta)) du$$

(5.6)

$$- \{ \int_0^1 \text{Cov}(h(u, \eta), \eta) du \}^2 / \text{Var}(\eta)]^{\frac{1}{2}} .$$

Using the same argument as in Lemma 3.1 of Holst (1972), we have that this quantity is maximized when

(5.7)
$$h(u, j) = \ell(u) \cdot j .$$

The results on the asymptotic distributions follow directly from Theorem 3.1 and Corollary 3.1 for the above special case.   □

From this result, it follows that the AMP test of level α is explicitly given by: Reject H_0 if

(5.8)
$$[\sum_{k=1}^m \ell(k/(m+1)(S_k - 1/\rho)] / [m(\int_0^1 \ell^2(u)du) (1+\rho)\rho^{-2}]^{\frac{1}{2}} > \lambda_\alpha$$

where λ_α is the upper α-percentile of the $N(0, 1)$ distribution. Also from Theorem 5.1, we find that the asymptotic power of this test in terms of the standard normal cdf is given by the expression

(5.9)
$$\Phi(-\lambda_\alpha + (\int_0^1 \ell^2(u)du/(1+\rho))^{\frac{1}{2}}) .$$

Furthermore it is easily seen from Theorem 3.1 that the Pitman Asymptotic Relative Efficiency (ARE) in using $h(u, j) = d(u) \cdot j$ instead of the optimal $h(u, j) = \ell(u) \cdot j$ is

-48-

$$(5.10) \quad e = (\int_0^1 d(u) \, \ell(u) du)^2 / \{\int_0^1 d^2(u) du - (\int_0^1 d(u) du)^2\} \{\int_0^1 \ell^2(u) du\}$$

5.A   Applications: Translation alternatives

We now consider some applications of the above results on non-symmetric tests.   First we shall look at the translation alternatives.   Let $X_1, \ldots, X_{m-1}$ be absolutely continuous i.i.d. random variables with distribution function $F(x)$. Let $Y_1, \ldots, Y_n$ be i.i.d. with d.f. $G(x)$.   We wish to test

$$H_0: \quad G(x) = F(x)$$

against the sequence of translation alternatives

$$(5.11) \quad A_m^{(1)}: \quad G(x) = G_m(x) = F(x - \theta/m^{\frac{1}{2}}) \; .$$

Let $f(x) = F'(x)$ be continuous. Then as $m \to \infty$,

$$(5.12) \quad L_m(u) = m^{\frac{1}{2}}[G_m(F^{-1}(u)) - u] \to -\theta f(F^{-1}(u)) = L(u), \quad \text{say.}$$

And if $f'(x)$ exists and is continuous except for at most finitely many x's then, at the continuity points of $f'(F^{-1}(u))$, we have

$$(5.13) \quad \ell_m(u) \to \ell(u) = -\theta \, f'(F^{-1}(u))/f(F^{-1}(u)) \; .$$

We now apply Theorem 5.1 to the following special cases to obtain the asymptotically optimal test statistics based on $\{S_k\}$.

-49-

Ex. 1   (Wilcoxon test)

   Consider the logistic distribution function

$$F(x) = 1/(1+e^{-x}), \quad -\infty < x < \infty .$$

In this case, we get

$$-f'(F^{-1}(u))/f(F^{-1}(u)) = 1 - 2u$$

and the regularity assumptions are all satisfied. Therefore the optimal

statistic from Theorem 5.1 is based on

$$(5.14) \qquad \sum_{k=1}^{m} (1 - 2k/(m+1)) S_k = 2 \sum_{1}^{m-1} R_k/(m+1) - (m-1)m/(m+1)$$
$$+ 2n/(m+1) - n$$

where $R_1, \ldots, R_{m-1}$ are the ranks of the X-values in the combined sample.
Observe that the Wilcoxon 2-sample test is also based on $\sum_{k=1}^{m-1} R_k$. The
optimality of the Wilcoxon statistic for the logistic distribution is well-

known. See e.g. Lehmann (1959), p. 238. Theorem 5.1 gives the asymp-

totic distribution of the Wilcoxon test as a special case.

Ex. 2   (A van der Waerden or normal score type test)

   For the normal d. f.

$$F(x) = \Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2} t^2)dt, \quad -\infty < x < \infty$$

we find

$$-f'(F^{-1}(u))/f(F^{-1}(u)) = \Phi^{-1}(u) .$$

It can be verified that the required regularity conditions of Theorem 3.1 are satisfied. Hence the AMP test is based on the statistic

$$(5.15) \qquad T = \sum_{k=1}^{m} \Phi^{-1}(k/(m+1)) S_k$$

from Theorem 5.1. Using the facts

$$\int_0^1 (\Phi^{-1}(u))^2 du = \int_{-\infty}^{\infty} x^2 \varphi(x) dx = 1 \quad \text{and} \quad \sum_{k=1}^{m} \Phi^{-1}(k/(m+1)) = 0 \;,$$

we have under the null hypothesis that

$$(5.16) \qquad \mathcal{L}(T/m^{\frac{1}{2}}) \to N(0, \, (1+\rho)/\rho^2) \;.$$

From Theorem 5.1, the asymptotic power for a one-sided test of level α is $\Phi(-\lambda_{\alpha} + \theta(1+\rho)^{-\frac{1}{2}})$, the same as that of the Student's t-test.

To find the Pitman efficiency of the Wilcoxon test with respect to the test based on (5.15) (which is optimal) in this case, we only need to calculate (5.10). Since

$$\int_0^1 (2u-1)\Phi^{-1}(u) du = \int_{-\infty}^{\infty} 2(\Phi(x)-1)x \, \varphi(x) dx = 2\int_{-\infty}^{\infty} (\varphi(x))^2 \, dx = \pi^{-\frac{1}{2}}$$

and $\int_0^1 (2u-1)^2 du = 1/3, \quad \int_0^1 (\Phi^{-1}(u))^2 du = 1,$ using formula (5.10) the ARE

of Wilcoxon test versus the normal scores type test

(5.17) $$e = 3/\pi .$$

The test statistic (5.15) has the same asymptotic properties as the

Fisher-Yates-Terry-Hoeffding and van der Waerden's rank tests.

Ex. 3 (A median test)

Consider the double exponential distribution with density

$$f(x) = \tfrac{1}{2} e^{-|x|}, \quad -\infty < x < \infty .$$

We find that the conditions of Theorem 3.1 are satisfied and

$$f'(F^{-1}(u))/f(F^{-1}(u)) = \begin{cases} 1 & \text{for } u < \tfrac{1}{2} \\ -1 & \text{for } u > \tfrac{1}{2} . \end{cases}$$

Hence the optimal test is based on the statistic

(5.18) $$\sum_{k=1}^{[m/2]} S_k - \sum_{k=[m/2]+1}^{m} S_k = 2 \sum_{k=1}^{[m/2]} S_k - n ,$$

where $\sum_{k=1}^{[m/2]} S_k$ is the number of Y's below the median of the X's . This

is different from the usual median test. We easily find that under the null

hypothesis,

(4.19) $$\mathfrak{L}([2 \sum_{k=1}^{[m/2]} S_k - n]/m^{\tfrac{1}{2}}) \to N(0, (1+\rho)/\rho^2) .$$

The ARE under normal alternatives is obtained from

$$-\int_0^{\frac{1}{2}} \Phi^{-1}(u)du + \int_{\frac{1}{2}}^1 \Phi^{-1}(u)dx = 2\int_{\frac{1}{2}}^1 \Phi^{-1}(u)du = (2/\pi)^{\frac{1}{2}}$$

which gives

(5.20) $$e = 2/\pi \ .$$

Ex. 4 For the Cauchy distribution, we find

(5.21) $$\ell(u) = 2\tan\pi(u-\tfrac{1}{2})/[1+\{\tan\pi(u-\tfrac{1}{2})\}^2] \ .$$

Notice that $\ell(0) = \ell(1) = 0$ and small weights are given to the S_k's in the tails. Heuristically this is appropriate since the Cauchy distribution has heavy tails and hence random fluctuations are large.

5.B Applications: Scale alternatives

Next we consider absolutely continuous positive random variables under scale alternatives. Let $X_1, \ldots, X_{m-1}$ be i.i.d. $F(x)$ and $Y_1, \ldots, Y_n$ be i.i.d. $G(y)$ with $F(0) = G(0) = 0$. We wish to test

(5.22) $$H_0: \ G(x) = F(x) \ ,$$

against the scale alternatives

(5.23) $$H_1^{(m)}: \ G(x) = G_m(x) = F(x(1+\theta/m^{\frac{1}{2}})) \ .$$

If the density $f(x) = F'(x)$ is continuous, then as $m \to \infty$,

(5.24) $$L_m(u) = m^{\frac{1}{2}}(G_m(F^{-1}(u))-u) \to L(u) = -\theta f(F^{-1}(u)) \cdot F^{-1}(u) \ .$$

-53-

And if f'(x) exists and is continuous except for finitely many points, then analogous to (5.13),

$$(5.25) \qquad \ell_m(u) \to \ell(u) = -\theta[1 + f'(F^{-1}(u)) \cdot F^{-1}(u)/f(F^{-1}(u))]$$

where f' exists. Optimal statistics based on $\{S_k\}$ can be derived just as in the case of translation alternatives.

Ex. 5. (Savage or exponential score test)

For the exponential distribution $F(x) = (1 - e^{-x})$ for $x > 0$ we find

$$(5.26) \qquad \ell(u) = -\theta(1 + \log(1 - u)) \quad .$$

The assumptions of Theorem 3.1 can be verified and hence an optimal statistic is given from Theorem 5.1, by

$$(5.27) \qquad T = \sum_{k=1}^{m} \log(1 - k/(m+1)) \, (S_k - 1/\rho) \quad .$$

Since $\int_0^1 (1 + \log(1 - u))^2 \, du = 1$ we get that

$$(5.28) \qquad \mathcal{L}(T/m^{\frac{1}{2}}) \to N(0, \, (1+\rho)/\rho^2)$$

-54-

and that the asymptotic power is

(5.29)
$$\Phi(-\lambda_\alpha + \theta(1+\rho)^{-\frac{1}{2}}) \ .$$

The ARE of Wilcoxon statistic relative to T in (5.27) above is $3/4$.

The statistic T is an approximation to the Savage statistic (see Lehmann

(1975) p. 103). The UMP test for the above situation is the test based on
$\sum_{k=1}^{m-1} X_k / \sum_{k=1}^{n} Y_k$ which has the same asymptotic power (5.29) as the statistic
T in (5.27).

Ex. 6. (Capon and Klotz test)

Consider the folded normal distribution with distribution function

(5.30)
$$F(x) = 2\Phi(x) - 1 \quad \text{for } x \geq 0 \ .$$

The assumptions of Theorem 3.1 can be verified and we find the optimal test

statistic using the function

$$\ell(u) = 1 - [\Phi^{-1}((1+u)/2)]^2 \ .$$

From Theorem 4.1, the statistic

(5.31)
$$T = \sum_{k=1}^{m} (\Phi^{-1}[(1+k/(m+1))/2])^2 S_k$$

has under the null hypothesis, the distribution given by

(5.32)
$$\mathcal{L}(T/m^{\frac{1}{2}}) \to N(0, \ 2(1+\rho)/\rho^2)$$

and the asymptotic power

-55-

$$\Phi\left[-\lambda_\alpha + \theta(2/(1+\rho))^{\frac{1}{2}}\right] \ .$$

Again this asymptotic power is the same as that of the conventional F-test.

6. Some further results and applications - the symmetric case

In this section, we consider the sequence of alternatives given by (4.2) and symmetric test statistics of the form

(6.1)
$$T_\nu^* = \sum_{k=1}^{m} h(S_k) \ .$$

As in Section 5, we first consider a result on the optimal choice of the function $h(\cdot)$.

Theorem 6.1. For the sequence of alternatives given by (4.2), satisfying the conditions of Theorem 4.1, the asymptotically most powerful (AMP) test is of the form: Reject H_0 when

(6.2)
$$\sum_{k=1}^{m} S_k(S_k - 1) > c \ .$$

Proof: From Theorem 4.1, it follows that the asymptotic power of a test of the form (6.1) is a maximum when the quantity A given in (4.11) is maximized. Observe that

(6.3) $$\mathrm{Cov}(\eta, \ \eta(\eta-1) - 4\eta/\rho) = 0$$

and

(6.4) $$\mathrm{Var}(h(\eta)) - \mathrm{Cov}^2(h(\eta), \eta)/\mathrm{Var}(\eta) = \mathrm{Var}(h(\eta) - \beta\eta)$$

where β is the usual linear regression coefficient

(6.5) $$\beta = \mathrm{Cov}(h(\eta), \eta)/\mathrm{Var}(\eta) \ .$$

-56-

Therefore we can rewrite

$$2\rho^{-2}(1+\rho)^2 \, A/\int_0^1 \ell^{*2}(u)du = \mathrm{Cov}(h(\eta)-\beta\eta,\ \eta(\eta-1)-4\eta/\rho)/[\mathrm{Var}(h(\eta)-\beta\eta)]^{\frac{1}{2}}$$

(6.6)

$$= \mathrm{Cor}\,(h(\eta)-\beta\eta,\ \eta(\eta-1)-4\eta/\rho)\cdot[\mathrm{Var}(\eta(\eta-1)-4\eta/\rho)]^{\frac{1}{2}}$$

$$\leq [\mathrm{Var}(\eta(\eta-1)-4\eta/\rho)]^{\frac{1}{2}} = 2\rho^{-2}(1+\rho)$$

with equality in (6.6) if and only if

$$h(\eta) - \beta\eta = a[\eta(\eta-1)-4\eta/\rho] + b$$

for some real numbers a and b . Thus A is maximized by $h(\eta) = \eta(\eta-1)$

and

(6.7)
$$\max_h A = \int_0^1 \ell^{*2}(u)du/(1+\rho) \ . \qquad \square$$

Using Theorem 4.1, we further have that under H_0 ,

(6.8)
$$\mathcal{L}([\sum_{k=1}^m S_k(S_k-1) - 2m/\rho^2]/m^{\frac{1}{2}}[2\rho^{-2}(1+\rho)]) \to N(0,1)$$

and that the asymptotic power for a test of level α is

$$\Phi[-\lambda_\alpha + (\int_0^1 \ell^{*2}(u)du)/(1+\rho)] \ .$$

Further, from the above proof we see that the ARE in using $h(S_k)$

instead of $S_k(S_k-1)$ is

(6.9)
$$e = \mathrm{Cor}^2(h(\eta) - \beta\eta,\ \eta(\eta-1)-4\eta/\rho) \ .$$

The statistic $\sum_{k=1}^m S_k^2$ which is equivalent to $\sum_{k=1}^m S_k(S_k-1)$ was proposed

by Dixon (1940). Blumenthal (1963) and Rao (1969) discuss the ARE of this test while Blum and Weiss (1957) consider the consistency properties. Blum and Weiss (1957) also show that the Dixon test is asymptotically LMP against "linear" alternatives with density $\{1 + c(y - \frac{1}{2})\}$, $0 \le y \le 1$ ($|c| \le 2$) but we have shown that Dixon test is indeed AMP against alternatives of the form (4.2).

For a nonnegative integer r, if we define

(6.10)
$$h(x) = \begin{cases} 1 & \text{for } x = r \\ 0 & \text{otherwise ,} \end{cases}$$

then

$$T^*_\nu = \sum_{k=1}^{m} h(S_k)$$

is the statistic $Q_m(r)$, the proportion of values among $\{S_k\}$ which are equal to r. This statistic has been discussed in Blum and Weiss (1957) from the point of consistency. Our results establish the asymptotic normality of $Q_m(r)$ under H_0 as well as under the sequence of alternatives (4.2). After some computations we find from Corollary 4.1 that under the null hypothesis

(6.11)
$$\mathcal{L}(m^{\frac{1}{2}} [Q_m(r) - \rho/(1+\rho)^{r+1}]) \to N(0, \sigma^2)$$

where

(6.12)
$$\sigma^2 = \{\rho/(1+\rho)^{r+1}\} [1 - (\rho/(1+\rho)^{r+1}) \{1 + (r-1/\rho)^2 (\rho^2/(1+\rho))\}] .$$

-58-

The Wald-Wolfowitz run test (1940) is related to $Q_m(0)$. Let U be the number of runs of X's and Y's in the combined sample. The hypothesis H_0 is rejected when U/m is too small. From the definition of $Q_m(r)$, it follows easily that

$$(6.13) \qquad |(U/m) - 2(n/m)(1 - Q_m(0))| \leq 1/m \ .$$

Thus the asymptotic distribution of U/m is the same as that of $2\rho(1 - Q_m(0))$ and we thus have, under H_0,

$$(6.14) \qquad \mathcal{L}(m^{\frac{1}{2}} [(U/m) - 2/(1+\rho)]) \rightarrow N(0, 4\rho/(1+\rho)^3) \ .$$

Therefore the ARE of the run-statistic against the Dixon's statistic is $\rho/(1+\rho)$ as has been shown in Rao (1969).

7. Further remarks and discussion

It is interesting to note that the theory developed in this paper gives tests based on $\{S_k\}$ which are asymptotically equivalent to the corresponding rank tests in all the known situations discussed in Section 5. For a unified approach to the theory of rank tests see Chernoff and Savage (1958) or Hajek and Sidak (1967). We conjecture that this property is true in general i.e., given any rank test, one can construct a test of the form (5.2) which has asymptotically the same null distribution and power. If this is the case, then the theory presented here seems to lead to much simpler test statistics as compared to the corresponding optimal rank tests. It is interesting to note that the optimal tests considered in Section 5 are linear in

-59-

$\{S_k\}$ while this is not the case with the ranks. Further relationships be-
tween these two groups of tests is under investigation. It may also be re-
marked that the theory presented here covers the asymptotic theory of many
tests that are not based on ranks as for instance, the run test and the medium
test.

The theorems presented here can also be applied to study similar test
statistics when the samples are censored. For instance, suppose that the
samples are censored at the right by $X'_{[(m-1)q]}$, the $[(m-1)q]^{th}$ order
statistic in the X-sample. Under the same assumptions as in Theorem 5.1,
we obtain in the same way that optimal test statistic is given by

$$(7.1) \qquad T = \sum_{k=1}^{[(m-1)q]} \ell(k/(m+1))(S_k - 1/\rho) \ .$$

Under H_0 ,

$$\mathscr{e}(T/m^{\frac{1}{2}}) \to N(0, [\int_0^q \ell^2(u)du - (\int_0^q \ell(u)du))^2] (\rho+1)/\rho^2)$$

and the asymptotic power is

$$(7.2) \qquad \Phi(-\lambda_\alpha + (\int_0^q \ell^2(u)du)/\{[\int_0^q \ell^2(u)du - (\int_0^q \ell(u)du)^2] (1+\rho)\}^{\frac{1}{2}}) \ .$$

On the otherhand censoring in rank theory can not be treated as simply as
this. See for instance Rao, Savage and Sobel (1960).

Results on the asymptotic theory of general statistics based on $\{S_k\}$
have been obtained, under the null hypothesis, by Holst (1976b) using a
different approach. However the approach used there does not seem to yield
the asymptotic theory under the alternatives, of the kind derived here.

-60-

Throughout the paper we formulated the alternatives in terms of the sample size m of the first sample. But a symmetric way of formulating the alternatives would be to express it in terms of $N = (m+n)$. We then have

$$N^{\frac{1}{2}}(G_N(F^{-1}(u)) - u) \to L(u) \quad (1+1/\rho)^{\frac{1}{2}} = L_1(u), \quad \text{say} .$$

In terms of $L_1'(u) = \ell_1(u) = \ell(u)(1+1/\rho)^{\frac{1}{2}}$, we see that the asymptotic power of this alternative can be expressed as

$$\Phi\left[-\lambda_\alpha + \left(\int_0^1 \ell_1^2(u)du/(1+\rho)(1+\rho^{-1})\right)^{\frac{1}{2}}\right] .$$

If $\ell_1(u)$ does not depend on ρ, then $\rho = 1$ maximizes the power i.e., in large samples, choosing $m \approx n$ or equal sample sizes is optimal.

REFERENCES

[1] Blum, J. R. and Weiss, L. (1957), Consistency of certain two-sample tests. *Ann. Math. Statist.* 28, 242-246.

[2] Blumenthal, S. (1963), The asymptotic normality of two test statistics associated with the two-sample problem. *Ann. Math. Statist.* 34, 1513-1523.

[3] Chernoff, H. and Savage, I. R. (1958), Asymptotic normality and efficiency of certain nonparametric statistics. *Ann. Math. Statist.* 29, 972-994.

[4] Courant, R. and Hilbert, D. (1962), *Methods of Mathematical Physics,* Vol. 2. Interscience, New York.

[5] Darling, D. A. (1953), On a class of problems related to the random division of an interval. *Ann. Math. Statist.* 24, 239-253.

[6] Dixon, W. J. (1940), A criterion for testing the hypothesis that two samples are from the same population. *Ann. Math. Statist.* 11, 199-204.

[7] Hájek, J. and Šidák, Z. (1967), *Theory of Rank Tests.* Academic press, New York.

[8] Holst, L. (1972), Asymptotic normality and efficiency for certain goodness of fit tests. *Biometrika* 59, 137-145.

[9] Holst, L. (1976a), On multinomial sums. *Math. Research Ctr.,* *Univ. of Wisconsin Tech. Report #1629.*

[10] Holst, L. (1976b), Some limit theorems for the indistinguishable ball problem with applications in nonparametrics. Math. Research Ctr., Univ. of Wisconsin Tech. Report.

[11] Le Cam, L. (1958), Un théorème sur la division d'un intervalle par des points pris au hasard. Publ. Inst. Statist. Univ. Paris 7, 7-16.

[12] Lehmann, E. L. (1959), Testing statistical hypotheses. John Wiley, New York.

[13] Lehmann, E. L. (1975), Nonparametric statistical methods based on ranks. Holden-Day, Inc., San Francisco.

[14] Mann, H. B. and Whitney, D. R. (1947), On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50-60.

[15] Pyke, R. (1965), Spacings. Jour. Roy. Statist. Soc., Ser. B 27, 395-449.

[16] Rao, C. R. (1973), Linear statistical inference and its applications. John Wiley, New York.

[17] Rao, J. S. (1969), Some contributions to the analysis of circular data. Unpublished Ph.D. thesis, Indian Statistical Inst., Calcutta.

[18] Rao, J. S. and Sethuraman, J. (1969), Pitman efficiencies of tests based on spacings. Tech. Report No. 11, Research and Training School, Indian Statistical Inst., Calcutta.

[19] Rao, J. S. and Sethuraman, J. (1975), Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. Ann. Statist. 3, 299-313.

[20] Rao, U. V. R., Savage, I. R. and Sobel, M. (1960), Contributions to the theory of rank order statistics: The two sample censored case. Ann. Math. Statist. 31, 415-426.

[21] Wald, A. and Wolfowitz, J. (1940), On a test whether two samples are from the same population. Ann. Math. Statist. 11, 147-162.

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
1655		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
ASYMPTOTIC THEORY FOR SOME FAMILIES OF TWO-SAMPLE NONPARAMETRIC STATISTICS.	Summary Report — no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Lars Holst and J. S. Rao	DAAG29-75-C-0024

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Mathematics Research Center, University of 610 Walnut Street Wisconsin Madison, Wisconsin 53706	

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709	July 1976
	13. NUMBER OF PAGES
	64

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
MRC-TSR-1655	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Two sample problems, Nonparametric methods, Sample spacings, Occupancy numbers, Limit distributions, Asymptotic efficiencies

$X(1), \ldots, X(M-1)$, and $Y(1), \ldots, Y(N)$ $X'(1) < or = \cdots < or = X'(m-1)$

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $X_1, \ldots, X_{m-1}$ and $Y_1, \ldots, Y_n$ be independent random samples from two continuous distribution functions F and G. We wish to test $H_0: F = G$. Let $X'_1 < \ldots < X'_{m-1}$ be the ordered X-observations. Denote by $S_k^{(k)}$ the number of Y-observations falling between $X'_{k-1}^{(k-1)}$ and $X'_k^{(k)}$. Asymptotic distribution theory and limiting efficiencies are studied for best statistics of the form $\sum h(S_k)$ symmetrically in $\{S_k\}$ and $\sum h_k(S_k)$ which are not symmetric in $\{S_k\}$, where $h(\cdot)$ and $\{h_k(\cdot)\}$ are real valued functions saitsfying some simple regularity conditions.

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE

sum of $h(S(k))$

the sets $S(k)$

22/200   the set $h(k(\cdot))$